

LUIS FERNANDO JUSSIANI

**DESEMPENHO DO MÉTODO DE LAGRANGEANO
AUMENTADO COM PENALIDADE QUADRÁTICA**

Curitiba

2004

LUIS FERNANDO JUSSIANI

**DESEMPENHO DO MÉTODO DE LAGRANGEANO
AUMENTADO COM PENALIDADE QUADRÁTICA**

Dissertação apresentada como requisito parcial à obtenção do grau de Mestre em Ciências, Curso de Pós-Graduação em Métodos Numéricos em Engenharia, concentração em Programação Matemática, Setores de Tecnologia e Ciências Exatas, Universidade Federal do Paraná.

Orientador : Prof. Luiz Carlos Matioli

Curitiba

2004

Termo de Aprovação

LUIS FERNANDO JUSSIANI

DESEMPENHO DO MÉTODO DE LAGRANGEANO AUMENTADO COM PENALIDADE QUADRÁTICA

Dissertação aprovada como requisito parcial para obtenção do grau de Mestre em Ciências, com área de concentração em Programação Matemática, no Programa de Pós-Graduação em Métodos Numéricos em Engenharia da Universidade Federal do Paraná, pela banca examinadora formada pelos professores:

Orientador:

Prof. Luiz Carlos Matioli, Dr. Eng.
Departamento de Matemática, UFPR

Examinadores:

Prof. Clóvis Caesar Gonzaga, DSc.
Departamento de Matemática, UFSC

Prof.^a Elizabeth Wegner Karas, Dr. Eng.
Departamento de Matemática, UFPR

Curitiba, 17 de dezembro de 2004

A Deus;
À minha esposa Deisy;
Ao meu filho Eduardo;
À minha família.

Agradecimentos

A realização deste trabalho só foi possível graças a colaboração de diversas pessoas. Algumas merecem ser destacadas por seu papel relevante e imprescindível. Agradeço em especial:

- Ao meu orientador, Professor LUIZ CARLOS MATIOLI, pela orientação brilhante na execução deste trabalho. Como pessoa um grande amigo, que sempre respeitou minhas limitações mostrando muita dedicação e humildade.
- Agradeço à minha esposa DEISY, pelo apoio e carinho recebidos durante a realização deste trabalho.
- Ao amigo MÁRCIO pelo incentivo na disciplina de programação não-linear.
- Aos amigos ALEX, EDSON, JULIANA e SACHIKO, pelos momentos de estudos e descontrações indispensáveis.
- Aos professores, funcionários e colegas do CESEC, pela amizade e dedicação.
- A todos do Programa de Pós-Graduação em Métodos Numéricos em Engenharia e às pessoas que direta ou indiretamente contribuíram para a realização deste trabalho.

*“Comece fazendo o que é necessário, depois o que é possível,
e de repente você estará fazendo o impossível.”*

São Francisco de Assis

Conteúdo

Lista de Figuras	vii
Lista de Tabelas	viii
Introdução	1
1 Conceitos Fundamentais	5
1.1 Notação	6
1.2 Vetores	7
1.3 Matrizes	8
1.4 Conjuntos	9
1.5 Funções Reais	10
1.6 Mínimos de Funções	12
1.7 Convexidade em \mathcal{R}^n	12
1.8 Elipses e Elipsóides	16
1.9 Função Conjugada	19
1.10 Quase-distância e Distância de Bregman	22
2 Lagrangeano Aumentado	25
2.1 Função de Penalidade Coerciva pela Direita	26
2.2 Conjugada de uma Penalidade da Família \mathcal{P} e suas Propriedades	26
2.3 Construindo Penalidades da Família \mathcal{P}	27
2.4 Método de Lagrangeano Aumentado	29
2.5 Ponto Proximal	32
2.6 Relação entre Lagrangeano Aumentado e Ponto Proximal	34
2.7 Relação entre Ponto Proximal e Região de Confiança	36
3 O Caso Linear	40
3.1 Programação Linear	40
3.2 A Penalidade Quadrática	43
3.3 Ponto Proximal Aplicado ao Problema Dual, com Quase-distância Quadrática	44
3.4 Lagrangeano Aumentado para o Problema Linear com Penalidade Quadrática	46
3.5 Afim Escala	48
3.6 Equivalência entre Afim Escala e Lagrangeano Aumentado com Penalidade Quadrática	49

4	Penalidade Quadrática	52
4.1	A Penalidade Quadrática Clássica	53
4.2	A Função Quadrática	54
4.3	Funções de Penalidade p Quadráticas	55
4.4	Geometria das Quase-distâncias dadas pelas Conjugadas das Penalidades Quadráticas.	58
4.5	Uma Heurística sobre a Atualização do Parâmetro de Penalidade	61
5	Implementação	68
5.1	Parâmetro de Penalidade	68
5.2	Algoritmo para o método de Região de Confiança	70
5.3	Algoritmos Implementados para o Método de Lagrangeano Aumentado .	72
6	Testes Numéricos	77
6.1	Resultados Numéricos	80
6.2	Resultados sobre Heurística da Atualização do Parâmetro de Penalidade .	91
	Considerações Finais	94
	Bibliografia	95

Lista de Figuras

1.1	Convexidade de Conjuntos	13
1.2	Elipse	16
1.3	Elementos da Elipse	17
1.4	Elipse da Equação Dada	18
1.5	Função Exponencial e sua Conjugada	19
1.6	Função Barreira Logarítmica e sua Conjugada	20
1.7	Quadrática e sua Conjugada	21
1.8	Distância de Bregman	23
3.1	Passo do Método de Ponto Proximal	45
3.2	Passo do Método de Ponto Proximal Forçando μ Ficar não Negativo	46
4.1	Função Quadrática Clássica	53
4.2	Quadrática e sua Conjugada	54
4.3	Curvas de Nível das Aproximações Quadráticas de Quase-distâncias	58
4.4	Direções Duais Geradas pelo Algoritmo de Lagrangeano Aumentado	59
4.5	Coordenada Mínima	60

Lista de Tabelas

6.1	Problemas do CUTE	78
6.2	Resultados Numéricos	81
6.3	Comparação entre Quadrática e M2b	89
6.4	Comparação entre Quadrática Tipo1 e M2b Tipo 2	90
6.5	Comparação entre as Quadráticas	91
6.6	Comparação das Penalidades Tipo 1 e 2 Quadrática com Heurística	92
6.7	Comparação entre Quadrática com Heurística e M2b	92
6.8	Comparação entre Quadrática com Heurística Tipo 1 e M2b Tipo 2	93

Resumo

Neste trabalho, serão utilizadas duas metodologias para construção de funções de penalização para algoritmos de Lagrangeano Aumentado, aplicados a problemas de programação convexa com restrições. Métodos de Lagrangeano Aumentado partem normalmente de funções de penalização $\theta : \mathcal{R} \rightarrow \mathcal{R}$, estritamente convexas e crescentes, que são combinadas com multiplicadores de Lagrange para compor termos de penalização com os formatos: $(y, \mu) \in \mathcal{R} \times \mathcal{R}_{++} \mapsto p(y, u) = \mu\theta(y)$ e $(y, \mu) \in \mathcal{R} \times \mathcal{R}_{++} \mapsto p(y, u) = \theta(\mu y)$. Propõe-se uma função de penalização θ a ser usada no algoritmo de Lagrangeano Aumentado, definida por $y \in \mathcal{R} \mapsto \theta(y) = \frac{1}{2}y^2 + y$, sendo θ estritamente convexa, porém não-crescente em todo o seu domínio. Neste caso, em que as penalidades são quadráticas, os multiplicadores gerados pelo algoritmo de Lagrangeano Aumentado podem ser negativos, pois a derivada da função não é crescente em todo o seu domínio. Este problema é contornado aumentando-se o parâmetro de penalidade, conforme relações mostradas no Capítulo 2, entre os métodos de Ponto Proximal e Região de Confiança. Implementam-se os algoritmos de Lagrangeano Aumentado para problemas com restrições de desigualdades, utilizando duas metodologias para construção das funções de penalidades quadrática e m2b. Os resultados numéricos obtidos em Matlab ilustram a eficiência da penalidade quadrática.

Abstract

In this work, two methodologies are used for constructing penalization functions of Augmented Lagrangian algorithms, solving convex programming problems with constraints. Augmented Lagrangian methods are usually built from strictly convex and increasing penalization functions $\theta : \mathcal{R} \rightarrow \mathcal{R}$, combined with Lagrange multipliers μ to compose penalization terms: $(y, \mu) \in \mathcal{R} \times \mathcal{R}_{++} \mapsto p(y, u) = \mu\theta(y)$ and $(y, \mu) \in \mathcal{R} \times \mathcal{R}_{++} \mapsto p(y, u) = \theta(\mu y)$. The penalization function θ , defined by $y \in \mathcal{R} \mapsto \theta(y) = \frac{1}{2}y^2 + y$, is θ strictly convex, but non-increasing in all its domain. In this case, the multipliers generated by the Augmented Lagrangian algorithm can be negative. Therefore the derivative of the function is not increasing in all its domain. This problem has been turned around by increasing the penalty parameter, according to relations shown in chapter 2, between the Proximal Point and Trust-Region methods. Augmented Lagrangian algorithms are implemented and tested for problems with inequality constraints, using the quadratic and m2b penalty functions. The numeric results obtained in Matlab illustrate the efficiency of the quadratic penalty.

Introdução

Neste Trabalho, vamos apresentar as motivações que levaram à escolha deste tema, os objetivos pretendidos e uma breve descrição do conteúdo dos demais capítulos desta dissertação. Encontra-se também indicada uma bibliografia básica que estará representada entre colchetes com a seguinte notação [N.^o].

Motivação

A programação matemática trata do estudo de problemas de otimização e o desenvolvimento de métodos para resolvê-los.

Nos últimos anos esse processo de otimização tem sido utilizado em diversas áreas da ciência. Em função dos avanços computacionais, as empresas têm buscado cada vez mais utilizar as ferramentas de otimização para maximizar lucros e minimizar custos, tornando, dessa forma, a otimização uma área de pesquisa bastante atraente. Como a maior parte dos problemas de otimização é resolvida por meio de computadores e os problemas gerados são cada vez maiores e difíceis de resolver, deve-se procurar algoritmos cada vez mais eficientes para resolvê-los. Neste trabalho são apresentados alguns dos algoritmos existentes na literatura.

O nosso trabalho está centrado no método de Lagrangeano Aumentado. Em geral, os métodos de Lagrangeano Aumentado são usados para resolver problemas de programação não linear com restrições. Os problemas com restrições a serem abordados são do tipo com restrições de desigualdades, ou seja, o problema restrito será da seguinte forma:

$$\begin{aligned}
& \text{Minimizar } f(x) \\
& \text{s.a. } g_i(x) \leq 0, \quad i = 1, 2, \dots, m \quad (P) \\
& x \in \mathcal{R}^n
\end{aligned}$$

sendo $f : \mathcal{R}^n \rightarrow \mathcal{R}$ a função objetivo, $g_i : \mathcal{R}^m \rightarrow \mathcal{R}$ para $i = 1, 2, \dots, m$ são funções convexas, próprias e fechadas.

A metodologia dos métodos de Lagrangeano Aumentado é bastante simples. Consiste em resolver o problema restrito através de uma sequência de subproblemas irrestritos. Os processos são iterativos, sendo que cada iteração consiste em resolver um subproblema irrestrito, que pode ser resolvido por qualquer método de minimização irrestrita (em nosso caso, Região de Confiança [19]), e atualização dos parâmetros (multiplicadores e parâmetro de penalização). Os subproblemas irrestritos resolvidos geram uma sequência, cujo nome é primal. Uma outra sequência gerada pelo método é a de multiplicadores, a qual é uma sequência dual. Assim, são geradas duas sequências: (x^k) a primal e (μ^k) a dual.

Tem-se um resultado importante entre primal e dual, o qual garante que, sob hipóteses adequadas, a sequência de multiplicadores gerada pelo algoritmo de Lagrangeano Aumentado aplicado ao problema primal é a mesma obtida por um certo algoritmo de Ponto Proximal aplicado ao problema dual.

Toda teoria de convergência da sequência de multiplicadores gerada pelos algoritmos de Lagrangeano Aumentado é desenvolvida no dual, via métodos proximais [20]. A parte de computação é feita toda no primal.

A seguir descreveremos alguns dos objetivos que pretendemos alcançar.

Principais objetivos do trabalho

Dentre os objetivos que se pretende alcançar ao longo deste trabalho destacam-se os seguintes:

- Estudo de metodologias para construção de funções de penalização para algoritmos de Lagrangeano Aumentado;

- Estudo de funções $\theta : \mathcal{R} \rightarrow \mathcal{R}$ que são combinadas com multiplicadores de Lagrange para compor termos de penalização;
- No caso específico de θ quadrática continuamos a pesquisa desenvolvida em [17], em que é fornecida uma metodologia diferente das usuais para os algoritmos de Lagrangeano Aumentado, sendo que em [17] os autores não implementaram o algoritmo com esta metodologia e aqui estamos propondo uma implementação;
- Implementação em Matlab dos algoritmos de Lagrangeano Aumentado desenvolvidos aqui no trabalho;
- Comparação dos resultados obtidos pelos algoritmos de Lagrangeano Aumentado, utilizando duas metodologias para construção das funções de penalidades θ quadrática e θ m2b [17];

A seguir descreveremos o que será tratado em cada um dos capítulos desta dissertação.

Escopo da dissertação

O presente trabalho está estruturado em seis capítulos. Abaixo será descrito de uma forma bastante sucinta como está organizado cada capítulo desta dissertação.

No *Capítulo 1* é feita uma apresentação das notações utilizadas ao longo do texto. Os tópicos abordados nesse capítulo são: Vetores, Matrizes, Funções reais de várias variáveis, Funções vetoriais, Conjuntos topológicos, Mínimos de funções, Convexidade, Elipse, Função conjugada e Distância de Bregman. Esses tópicos formarão a parte teórica básica para o restante dos capítulos.

No *Capítulo 2* são descritos os conceitos envolvidos nos métodos de Lagrangeano Aumentado, Ponto Proximal e Região de Confiança. O principal resultado desse capítulo é o de equivalência das seqüências duais geradas pelos métodos de Lagrangeano Aumentado e de Ponto Proximal. Outra equivalência importante tratada nesse capítulo é a entre métodos de Ponto Proximal e de Região de Confiança.

No *Capítulo 3* serão apresentados os conceitos de programação linear. Algoritmo de Ponto Proximal aplicado ao problema dual com quase-distância quadrática, o algorit-

mo de Lagrangeano Aumentado para o problema linear com penalidade quadrática, e o algoritmo Afim Escala. O resultado principal do capítulo é que as direções duais geradas pelo algoritmo de Lagrangeano Aumentado, aplicado ao problema linear com penalidade quadrática são equivalentes às geradas pelo algoritmo Afim Escala aplicado ao dual do problema linear.

No *Capítulo 4* propõe-se uma metodologia de uso de função de penalização combinada com multiplicadores de Lagrange para o método de Lagrangeano Aumentado. No caso, em que as penalidades são quadráticas, os multiplicadores gerados pelo algoritmo de Lagrangeano Aumentado podem ser negativos, pois a derivada da função não é crescente em toda a reta. No final desse capítulo, apresenta-se uma heurística em que é possível contornar o problema de multiplicadores negativos encontrando um valor para o parâmetro r que torne-os positivos.

No *Capítulo 5* serão tratadas questões referentes à implementação dos algoritmos de Lagrangeano Aumentado desenvolvidos nesta dissertação, além disso, serão abordadas as questões da atualização do parâmetro de penalidade.

No *Capítulo 6* serão apresentados os testes numéricos do algoritmo de Lagrangeano Aumentado com a metodologia proposta. Aplica-se também o algoritmo com a penalidade m2b [17] e faz-se uma análise comparativa dos resultados encontrados entre as duas metodologias.

Finalizando, serão apresentadas as conclusões obtidas com o desenvolvimento do trabalho e algumas propostas para outros possíveis trabalhos nessa área. E, por final, as referências bibliográficas.

Capítulo 1

Conceitos Fundamentais

Neste capítulo, apresentaremos algumas ferramentas e notações que serão utilizadas nos demais capítulos desta dissertação.

Serão introduzidos conceitos matemáticos básicos como: vetores, matrizes, funções reais de várias variáveis, funções vetoriais e de conjuntos.

Apresentaremos uma seção referente a mínimos de função. Cada um dos conceitos de mínimos será abordado de forma rápida e objetiva, para posteriormente serem utilizados no decorrer do trabalho.

Abordaremos algumas definições de conjunto convexo, bem como de funções convexas, sendo muito importante em otimização, pois estão relacionadas com o conceito de mínimo global.

Apresentaremos, também, uma seção sobre elipses e elipsóides, tendo como objetivo fornecer subsídios para o método Afim Escala discutido nos Capítulos 3 e 4.

Em seguida, exporemos a definição de função conjugada, e essa será muito importante quando formos discutir, mais adiante, as relações entre os métodos de Lagrangeano Aumentado e de Ponto Proximal.

Finalizando, apresentaremos a definição de distância de Bregman e algumas de suas propriedades mais usuais. A distância de Bregman será utilizada no próximo capítulo em Ponto Proximal.

1.1 Notação

Listamos, a seguir, algumas notações que serão utilizadas com mais frequência.

1. $x \in \mathcal{R}^n$ é vetor coluna em \mathcal{R}^n ;
2. $x^T = (x_1, x_2, \dots, x_n)$ - vetor transposto;
3. $\langle x, y \rangle$ - produto escalar (usual em \mathcal{R}^n);
4. $e = (1, 1, \dots, 1)^T$ - vetor unitário em \mathcal{R}^n ;
5. $\|\cdot\|$ - norma Euclidiana;
6. $\text{dom}f$ - domínio efetivo da função f ;
7. $\text{intdom}f$ - interior do domínio efetivo da função f ;
8. $\text{ri}X$ - interior relativo do conjunto X ;
9. $\text{int}(X)$ - interior relativo do conjunto X ;
10. \overline{X} - fecho do conjunto X ;
11. $f'(x, d)$ - derivada direcional de f no ponto x e na direção d ;
12. $\partial f(\cdot)$ - subdiferencial da função f ;
13. ∂X - fronteira do conjunto X ;
14. $f'_\infty(\cdot)$ - função de recessão da função $f(\cdot)$;
15. $\mathcal{R}_{++}^m = \{x \in \mathcal{R}^m : x_i > 0, i = 1, \dots, m\}$;
16. $\mathcal{R}_+^m = \{x \in \mathcal{R}^m : x_i \geq 0, i = 1, \dots, m\}$;
17. Usamos ainda uma notação simplicada para seqüências, escrevendo (x^k) ao invés de $\{x^k\}_{k \in \mathcal{N}}$.

1.2 Vetores

Trabalharemos com vetores colunas, denotados por letras minúsculas. Um vetor $x \in \mathcal{R}^n$ será indicado por

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \quad (1.1)$$

onde $x_i \in \mathcal{R}$ para $i = 1, 2, \dots, n$ são as i -ésimas componentes do vetor x em relação a uma base pré-definida.

Denotamos o vetor $x^T = (x_1, x_2, \dots, x_n)$ como sendo o vetor transposto de x , dado na relação (1.1).

O **produto escalar** entre dois vetores x e y em \mathcal{R}^n é definido por

$$\langle x, y \rangle = \sum_{i=1}^n x_i \cdot y_i = x^T \cdot y. \quad (1.2)$$

O **vetor unitário “e”** denota o vetor de uns, isto é,

$$e = (1, \dots, 1)^T. \quad (1.3)$$

A **norma euclidiana** de um vetor $x \in \mathcal{R}^n$ é definida por

$$\|x\| = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2} = \sqrt{x^T \cdot x}. \quad (1.4)$$

A **distância** entre dois pontos x e y em \mathcal{R}^n é definida por

$$D(x, y) = \|y - x\|. \quad (1.5)$$

1.3 Matrizes

As matrizes são denotadas por letras maiúsculas. Uma matriz em $\mathcal{R}^{m \times n}$ será indicada por

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} = [a_{ij}] \quad (1.6)$$

onde m é o número de linhas e n o número de colunas da mesma.

Apresentamos a seguir algumas notações de matrizes que serão utilizadas ao longo do texto.

- i) A^T : **Matriz Transposta** de A, onde $A^T = [a_{ji}]$;
- ii) I : **Matriz Identidade**, é uma matriz quadrada com uns em sua diagonal principal, isto é,

$$\begin{cases} a_{ij} = 1 & \text{se } i = j \\ a_{ij} = 0 & \text{se } i \neq j \end{cases}; \quad (1.7)$$

- iii) A^{-1} : **Matriz Inversa** de A;
- iv) Uma matriz quadrada é **Simétrica** se : $a_{ij} = a_{ji}$ para todo i e j ;
- v) Uma matriz A real simétrica é **semidefinida positiva** se, e somente se

$$x^T \cdot A \cdot x \geq 0 \quad \forall x \in \mathcal{R}^n; \quad (1.8)$$

- vi) Uma matriz A real simétrica é **definida positiva** se, e somente se

$$x^T \cdot A \cdot x > 0 \quad \forall x \in \mathcal{R}^n \text{ tal que } x \neq 0. \quad (1.9)$$

1.4 Conjuntos

Nesta seção, abordaremos algumas definições topológicas elementares referentes a subconjuntos de \mathcal{R}^n , visando estabelecer uma base para desenvolver os capítulos seguintes. Algumas destas definições são baseadas em [15] e [18].

Definição 1.4.1 Define-se a **bola aberta** de raio Δ centrada em $x \in \mathcal{R}^n$ como

$$\beta_{\Delta}(x) = \{y \in \mathcal{R}^n : \|x - y\| < \Delta\}. \quad (1.10)$$

Definição 1.4.2 Seja $\mathcal{S} \subset \mathcal{R}^n$ e $x \in \mathcal{S}$. Então x é um **ponto interior** a \mathcal{S} se existe uma bola $\beta_{\Delta}(x)$ inteiramente contida em \mathcal{S} . O conjunto de todos os pontos interiores de \mathcal{S} será denotado $\text{int}(\mathcal{S})$.

Definição 1.4.3 O subconjunto $\mathcal{S} \subset \mathcal{R}^n$ é **aberto** se $\mathcal{S} = \text{int}(\mathcal{S})$.

Definição 1.4.4 Seja $\mathcal{S} \subset \mathcal{R}^n$ e $x \in \mathcal{R}^n$. Então x é um **ponto fronteira** de \mathcal{S} se toda a bola $\beta_{\Delta}(x)$ (1.10) contém pontos de \mathcal{S} e pontos fora de \mathcal{S} . O conjunto de todos os pontos de fronteira de \mathcal{S} será representado por $\partial\mathcal{S}$.

Definição 1.4.5 Seja $\mathcal{S} \subset \mathcal{R}^n$, o **fecho** (closure) de \mathcal{S} , denotado por $\overline{\mathcal{S}}$ ou $\text{cl}(\mathcal{S})$, é a união de \mathcal{S} com a sua fronteira, isto é, $\overline{\mathcal{S}} = \mathcal{S} \cup \partial\mathcal{S}$.

Definição 1.4.6 O subconjunto $\mathcal{S} \subset \mathcal{R}^n$ é **fechado** se coincide com o fecho de \mathcal{S} , isto é, se $\mathcal{S} = \overline{\mathcal{S}}$.

Definição 1.4.7 Seja $\mathcal{S} \subset \mathcal{R}^n$. Então \mathcal{S} é um **conjunto limitado**, se existe um escalar $M > 0$ tal que $\|x\| \leq M$ para todo $x \in \mathcal{S}$.

Definição 1.4.8 Um subconjunto $\mathcal{S} \subset \mathcal{R}^n$ é **compacto**, se \mathcal{S} é ao mesmo tempo limitado e fechado.

Teorema 1.4.9 (Bolzano-Weierstrass). Toda seqüência limitada em \mathcal{R}^n possui uma subseqüência convergente.

Prova. Ver [15] ■

Em virtude do teorema (1.4.9), um conjunto $\mathcal{S} \subset \mathcal{R}^n$ é compacto se, e somente se, toda seqüência de pontos $x \in \mathcal{S}$ possui uma subseqüência que converge para um ponto de \mathcal{S} .

1.5 Funções Reais

As funções serão definidas no conjunto $\mathcal{S} \subseteq \mathcal{R}^n$ e assumindo valores em \mathcal{R} , isto é, $f : \mathcal{S} \subseteq \mathcal{R}^n \rightarrow \mathcal{R}$.

A **derivada direcional** de f no ponto x^0 , ao longo da direção $d \in \mathcal{R}^n$, é definida como

$$f'(x^0, d) = \lim_{\alpha \rightarrow 0} \frac{f(x^0 + \alpha d) - f(x^0)}{\alpha}. \quad (1.11)$$

O vetor das derivadas parciais de $f(x)$ em um ponto x , é denominado de vetor **Gradiente**, e indicado por

$$\nabla f(x) = \left[\frac{\partial f(x)}{\partial x_i} \right] = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{bmatrix}. \quad (1.12)$$

A matriz das derivadas parciais de segunda ordem de $f(x)$ em um ponto x , é denominado de **matriz Hessiana**, e indicada por

$$H(x) = \nabla^2 f(x) = \left[\frac{\partial^2 f(x)}{\partial x_i \partial x_j} \right] = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_n \partial x_n} \end{bmatrix}. \quad (1.13)$$

O conjunto de várias funções a valores reais g_1, g_2, \dots, g_m em \mathcal{R}^n , pode ser visto como uma função vetorial $g(x)$. Essa função determina o vetor $g(x)$ definido por $g : \mathcal{R}^n \rightarrow \mathcal{R}^m$, ou seja,

$$g(x)^T = [g_1(x), g_2(x), \dots, g_m(x)]. \quad (1.14)$$

Supondo que g seja contínua com derivadas parciais contínuas. A matriz J das derivadas parciais de g , denomina-se **matriz Jacobiana** e será indicada por

$$J(x) = \nabla g(x) = \left[\frac{\partial g_i(x)}{\partial x_j} \right] = \begin{bmatrix} \frac{\partial g_1(x)}{\partial x_1} & \frac{\partial g_1(x)}{\partial x_2} & \dots & \frac{\partial g_1(x)}{\partial x_n} \\ \frac{\partial g_2(x)}{\partial x_1} & \frac{\partial g_2(x)}{\partial x_2} & \dots & \frac{\partial g_2(x)}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_m(x)}{\partial x_1} & \frac{\partial g_m(x)}{\partial x_2} & \dots & \frac{\partial g_m(x)}{\partial x_n} \end{bmatrix}. \quad (1.15)$$

A seguir apresentaremos um resultado fundamental e básico em análise. A prova deste é baseado em [18].

Teorema 1.5.1 *Se f é uma função real e contínua, definida em um conjunto compacto $\mathcal{S} \subset \mathcal{R}^n$ (\mathcal{S} é fechado e limitado), então o problema de otimização*

$$\begin{aligned} & \text{Minimizar } f(x) \\ & x \in \mathcal{S} \end{aligned}$$

tem solução ótima $x^ \in \mathcal{S}$.*

Prova. *Seja $m = \inf_{x \in \mathcal{S}} \{f(x)\}$ (assim $\forall x \in \mathcal{S} : m \leq f(x)$). Então existe uma seqüência (x^k) de elementos de \mathcal{S} tal que $f(x^k) \rightarrow m$.*

Sendo \mathcal{S} compacto, possui uma subsequência $\{x^l\}_{l \in L}$ ($L \subset \mathcal{N}$) que converge para $x^ \in \mathcal{S}$.*

Da continuidade de f , tem-se $f(x^l) \rightarrow f(x^)$ e*

$$m = \lim_{k \rightarrow \infty} f(x^k) = \lim_{l \rightarrow \infty} f(x^l) = f(x^*)$$

Como $f(x^) > -\infty$, tem-se $m > -\infty$ e $\forall x \in \mathcal{S} : f(x^*) = m \leq f(x)$, portanto $x^* \in \mathcal{S}$ é uma solução ótima do problema. ■*

O resultado do teorema (1.5.1) é fundamental, pois trata da existência de uma solução ótima para um problema de otimização.

1.6 Mínimos de Funções

No processo de minimização de funções, pode-se obter como solução dos problemas diversos tipos de minimizadores.

Considere o problema

$$\begin{aligned} & \text{Minimizar } f(x) \\ & \text{s.a : } x \in \mathcal{S} \end{aligned} \tag{1.16}$$

onde $f : \mathcal{R}^n \rightarrow \mathcal{R}$ e $\mathcal{S} \subset \mathcal{R}^n$. \mathcal{S} é chamado conjunto viável e (1.16) é a forma genérica dos problemas de programação não-linear.

Definição 1.6.1 (*Mínimo local*) Um ponto $x^* \in \mathcal{S}$ é um minimizador local de f se, e somente se, existe $\varepsilon > 0$ tal que $f(x) \geq f(x^*)$ para todo $x \in \mathcal{S}$ tal que $\|x - x^*\| < \varepsilon$.

Definição 1.6.2 (*Mínimo local estrito*) Um ponto $x^* \in \mathcal{S}$ é um minimizador local estrito de f se, e somente se, existe $\varepsilon > 0$ tal que $f(x) > f(x^*)$ para todo $x \in \mathcal{S}$ tal que $x \neq x^*$ e $\|x - x^*\| < \varepsilon$.

Definição 1.6.3 (*Mínimo global*) Um ponto $x^* \in \mathcal{S}$ é um minimizador global de f se, e somente se, $f(x) \geq f(x^*)$ para todo $x \in \mathcal{S}$.

Definição 1.6.4 (*Mínimo global estrito*) Um ponto $x^* \in \mathcal{S}$ é um minimizador global estrito de f se, e somente se, $f(x) > f(x^*)$ para todo $x \in \mathcal{S}$ tal que $x \neq x^*$.

De forma análoga, define-se os maximizadores locais e globais. Observando que *maximizar* f é equivalente a *minimizar* $-f$.

1.7 Convexidade em \mathcal{R}^n

A definição de problema convexo é muito importante em otimização, pois está relacionada com o conceito de mínimo global. Algumas destas propriedades estão demonstradas nos livros [7], [8], [11], [12] e [18], nos quais estamos nos baseando.

Definição 1.7.1 Um conjunto $\mathcal{S} \subset \mathcal{R}^n$ é um conjunto convexo, se o segmento de linha que liga quaisquer dois pontos de \mathcal{S} estiver também em \mathcal{S} . Equivalentemente, o conjunto $\mathcal{S} \subset \mathcal{R}^n$ é convexo se, e somente se, para todo $x, y \in \mathcal{S}$ e $\lambda \in \mathcal{R}$ com $0 \leq \lambda \leq 1$ tem-se

$$(1 - \lambda)x + \lambda y \in \mathcal{S}. \quad (1.17)$$

Geometricamente quer dizer que se escolhermos dois pontos quaisquer no interior do domínio e os unirmos por um segmento de reta, e se todos os pontos desse segmento estiverem no interior do domínio, esse será considerado convexo. Caso contrário, será um domínio não-convexo. A figura.(1.1) ilustra esse conceito.

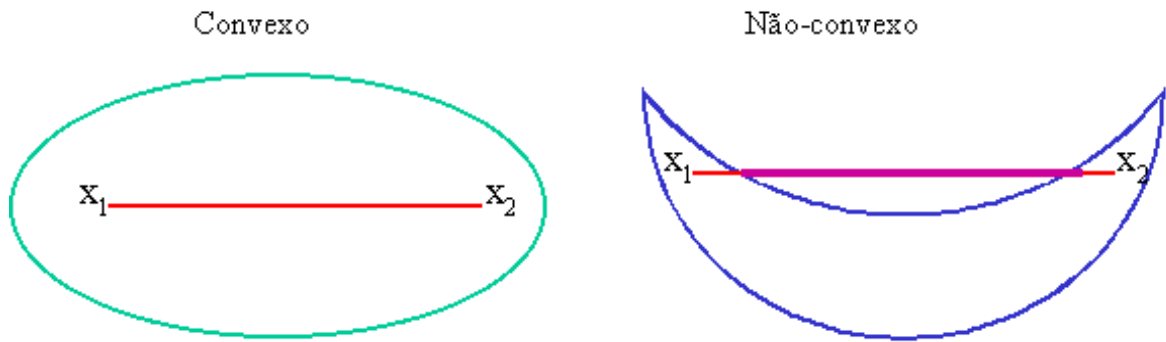


Figura 1.1: Convexidade de Conjuntos

Teorema 1.7.2 Seja $\mathcal{S}_i = 1, \dots, n$, conjuntos convexos. Então o conjunto

$$\mathcal{S} = \bigcap_{i=1}^n \mathcal{S}_i \quad (1.18)$$

é também convexo.

Prova. Ver [8] ■

Definição 1.7.3 Uma função $f : \mathcal{R}^n \rightarrow \mathcal{R} \cup \{\infty\}$ definida em um conjunto $\mathcal{S} \subset \mathcal{R}^n$ é convexa, se para todo $x, y \in \mathcal{S}$ e $\lambda \in \mathcal{R}$, $0 \leq \lambda \leq 1$

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y). \quad (1.19)$$

Se a desigualdade em (1.19) é estrita a função é chamada de estritamente convexa.

Teorema 1.7.4 *Se $f \in C^1$ então (i) e (ii) abaixo são equivalentes, e se $f \in C^2$ então (i), (ii) e (iii) abaixo são equivalentes.*

$$\begin{aligned} & i) \quad f \text{ é convexa}; \\ & ii) \quad \forall x, y \in \mathcal{R}^n : f(y) \geq f(x) + \nabla f^T(x) \cdot (y - x); \\ & iii) \quad \nabla^2 f(x) \geq 0, \forall x \in \mathcal{R}^n. \end{aligned} \tag{1.20}$$

Teorema 1.7.5 *Seja $f : \mathcal{S} \rightarrow \mathcal{R}$ função convexa e $\alpha \in \mathcal{R}$. Então o conjunto de nível $\{x : f(x) \leq \alpha\}$ é convexo.*

Ver [16].

Definição 1.7.6 *O domínio efetivo de uma função convexa é o conjunto não vazio*

$$\text{dom} f = \{x \in \mathcal{R}^n : f(x) < \infty\}. \tag{1.21}$$

Definição 1.7.7 *Dizemos que uma função convexa $f : \mathcal{R}^n \rightarrow \mathcal{R} \cup \{\infty\}$ é **própria** se é finita em pelo menos um ponto do seu domínio.*

Definição 1.7.8 *Considere $f : \mathcal{R}^n \rightarrow \mathcal{R} \cup \{\infty\}$ convexa e própria. O vetor $s \in \mathcal{R}^n$ é um **subgradiente** de f no ponto $x \in \mathcal{R}^n$ se satisfaz:*

$$\forall y \in \mathcal{R}^n, \quad f(y) \geq f(x) + s^T(y - x).$$

O conjunto de todos os subgradientes de f no ponto x é chamado de subdiferencial de f em x , e é denotado por

$$\partial f(x) = \{s \in \mathcal{R}^n : f(y) \geq f(x) + s^T(y - x), \forall y \in \mathcal{R}^n\}.$$

Definição 1.7.9 *Considere $x \in \mathcal{R}^n \mapsto g(x) \in \mathcal{R} \cup \{\infty\}$ uma função convexa, própria, fechada e $\mathcal{S} \subset \mathcal{R}^n$ aberto tal que $\mathcal{S} \subset \text{dom } g$. Dizemos que g é **coerciva na fronteira** de \mathcal{S} se, dados $x \in \partial \mathcal{S}$, $y \in \mathcal{S}$ e $h = y - x$ ($x \neq y$), então*

$$\lim_{\substack{z \rightarrow x \\ z \in \mathcal{S}}} g'(z, h) = -\infty$$

onde $g'(z, h)$ é a derivada direcional de g em z na direção h .

Definição 1.7.10 O *epígrafo* da função f é um conjunto não vazio, denotado por $\text{epi}(f)$, e definido por

$$\text{epi}(f) = \{(x, r) \in \mathcal{R}^n \times \mathcal{R} : r \geq f(x)\} \subset \mathcal{R}^{n+1},$$

e, para cada $r \in \mathcal{R}$, o conjunto de nível r de f é definido como

$$S_r = \{x \in \mathcal{R}^n : f(x) \leq r\}.$$

Proposição 1.7.11 A função f é convexa se, e somente se, $\text{epi}(f)$ é um conjunto convexo.

Considere o problema

$$\begin{aligned} & \text{Minimizar} \quad f(x) \\ & \text{s.a.} \quad g_i(x) \leq 0, \quad i = 1, \dots, m \\ & \quad x \in \mathcal{S} \subset \mathcal{R}^n \end{aligned} \tag{1.22}$$

O problema (1.22) é um problema de programação convexa (ou simplesmente problema convexo) se

$$\left\{ \begin{array}{l} f \text{ é convexa;} \\ \text{as funções } g_i (i = 1, \dots, m) \text{ são convexas;} \\ \mathcal{S} \subset \mathcal{R}^n \text{ é convexo.} \end{array} \right.$$

A grande vantagem de um problema ser convexo é que apresenta um mínimo global. A seguir apresentamos um teorema que é fundamental em programação convexa.

Teorema 1.7.12 Em problemas de programação convexa todo ótimo local é um ótimo global.

Prova. Ver [18] ■

1.8 Elipses e Elipsóides

Nesta seção apresentamos a definição de elipse e elipsóide. Não entraremos em muitos detalhes, pois pretendemos apenas lembrar ao leitor as relações de elipses e de elipsóides.

Definição 1.8.1 Uma **Elipse** é o conjunto de todos os pontos de um plano cuja soma das distâncias a dois pontos fixos (focos) é constante.

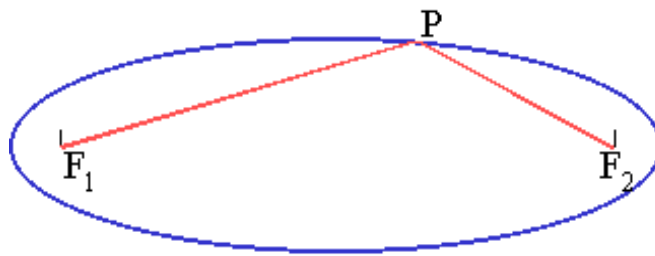


Figura 1.2: Elipse

A equação de uma elipse que tem centro na origem eixos paralelos aos eixos coordenados, pode ser escrita como

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1. \quad (1.23)$$

Com $a^2 > b^2$, a elipse tem os vértices $(\pm a, 0)$ e $(0, \pm b)$, e os focos são $(\pm c, 0)$, sendo $c^2 = a^2 - b^2$. Da equação dada na relação (1.23) tem-se a figura (1.3)

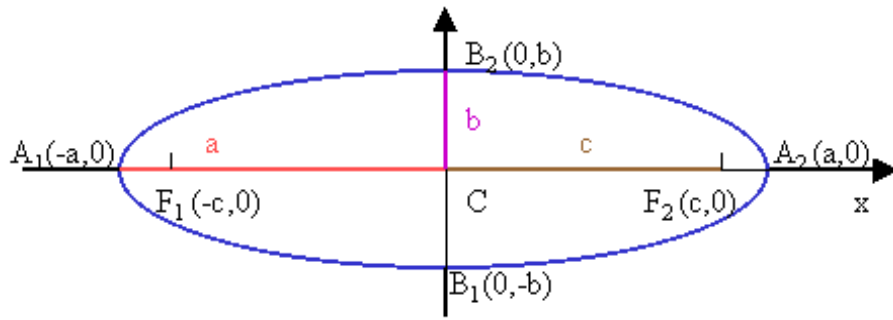


Figura 1.3: Elementos da Elipse

A equação de uma elipse de centro (h, k) com eixos paralelos aos eixos coordenados, pode ser escrita como

$$\frac{(x - h)^2}{a^2} + \frac{(y - k)^2}{b^2} = 1. \quad (1.24)$$

Se $h = 0$ e $k = 0$ a equação (1.24) é a equação (1.23).

Desenvolvendo os quadrados indicados na equação dada na relação (1.24) e simplificando termos, obtém-se uma equação da forma

$$Ax^2 + By^2 + Cx + Dy + E = 0 \quad (1.25)$$

onde os coeficientes são números reais e A e B são ambos positivos. Reciprocamente, partindo da equação (1.25) e completando quadrado, pode-se obter uma forma que evidencie o centro da elipse e o comprimento do maior e menor eixo.

Exemplo 1.8.2 *Seja equação*

$$\left(\frac{x}{3} - 1\right)^2 + \left(\frac{y}{2} - 1\right)^2 = 1. \quad (1.26)$$

A figura (1.4) a seguir representa o gráfico da elipse dada na relação (1.26)

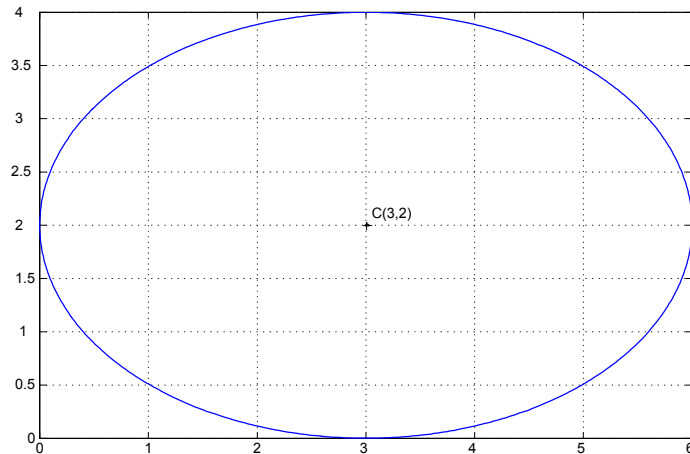


Figura 1.4: Elipse da Equação Dada

A equação (1.26) pode ser escrita como

$$\frac{(x-3)^2}{3^2} + \frac{(y-2)^2}{2^2} = 1. \quad (1.27)$$

Observe que (1.27) tem centro $C(3,2)$, $a=3$, $b=2$ e $c=\sqrt{5}$, excentricidade $e=\frac{c}{a}=\frac{\sqrt{5}}{3}$ e focos $F_1(3-\sqrt{5},2)$ e $F_2(3+\sqrt{5},2)$. Desenvolvendo os quadrados de (1.27) tem-se

$$4x^2 + 9y^2 - 24x - 36y + 36 = 0.$$

O **elipsóide** de centro na origem é a superfície representada por

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1 \quad (1.28)$$

onde a , b , c são números reais positivos.

Se o centro do elipsóide é o ponto (h, k, l) e seus eixos forem paralelos aos eixos coordenados, a equação (1.28) assume a forma

$$\frac{(x-h)^2}{a^2} + \frac{(y-k)^2}{b^2} + \frac{(z-l)^2}{c^2} = 1$$

obtida através de uma translação de eixos.

1.9 Função Conjugada

A função conjugada é definida a partir de uma função $f : \mathcal{R}^n \rightarrow \mathcal{R} \cup \{+\infty\}$.

Definição 1.9.1 Considere $f : \mathcal{R}^n \rightarrow \mathcal{R} \cup \{+\infty\}$ uma função convexa. A convexa conjugada de f é a função f^* dada por:

$$s \in \mathcal{R}^n \mapsto f^*(s) = \sup\{x^T s - f(x), x \in \mathcal{R}^n\} \quad (1.29)$$

Teorema 1.9.2 A conjugada f^{**} de f^* é a própria f e f^* é fechada, própria e convexa se, e somente se, f é própria.

Prova. Ver [21] ■

Exemplo 1.9.3 Considere a função $f : \mathcal{R} \rightarrow \mathcal{R} \cup \{+\infty\}$ definida por

$$f(x) = e^x - 1,$$

f tem como conjugada f^* dada por

$$s \in \mathcal{R} \mapsto f^*(s) = \begin{cases} s \ln s - s + 1, & \text{se } s \geq 0 \\ +\infty & , \text{ se } s < 0 \end{cases}. \quad (1.30)$$

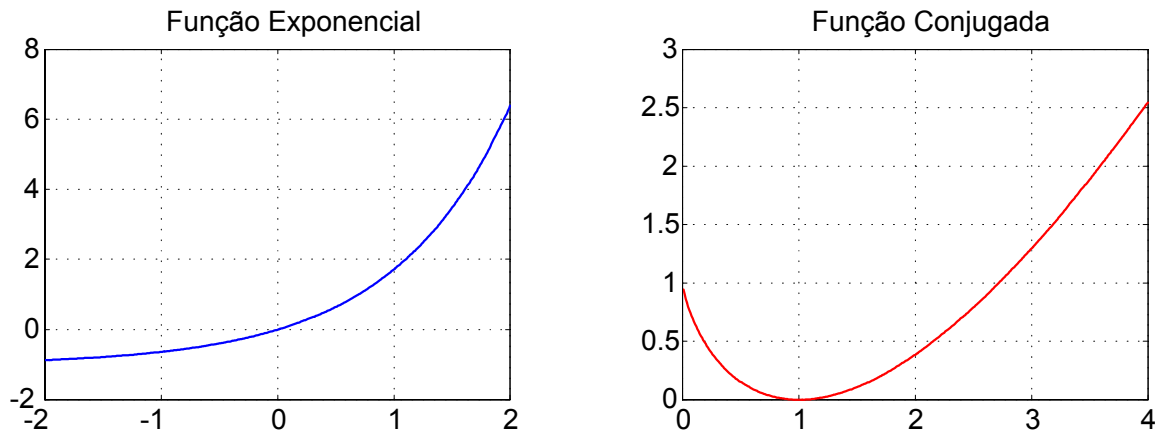


Figura 1.5: Função Exponencial e sua Conjugada

A figura (1.5) mostra o gráfico da função f e da sua conjugada f^* . A função conjugada da figura (1.5) é coerciva na fronteira do ortante positivo.

Exemplo 1.9.4 Considere a função $f : \mathcal{R} \rightarrow \mathcal{R} \cup \{+\infty\}$ definida por

$$f(x) = \begin{cases} -\ln(1-x), & \text{se } x < 1 \\ +\infty & , \text{ se } x \geq 1 \end{cases},$$

f tem como conjugada f^* dada por

$$s \in \mathcal{R} \mapsto f^*(s) = \begin{cases} s - \ln s - 1, & \text{se } s > 0 \\ +\infty & , \text{ se } s \leq 0 \end{cases}. \quad (1.31)$$

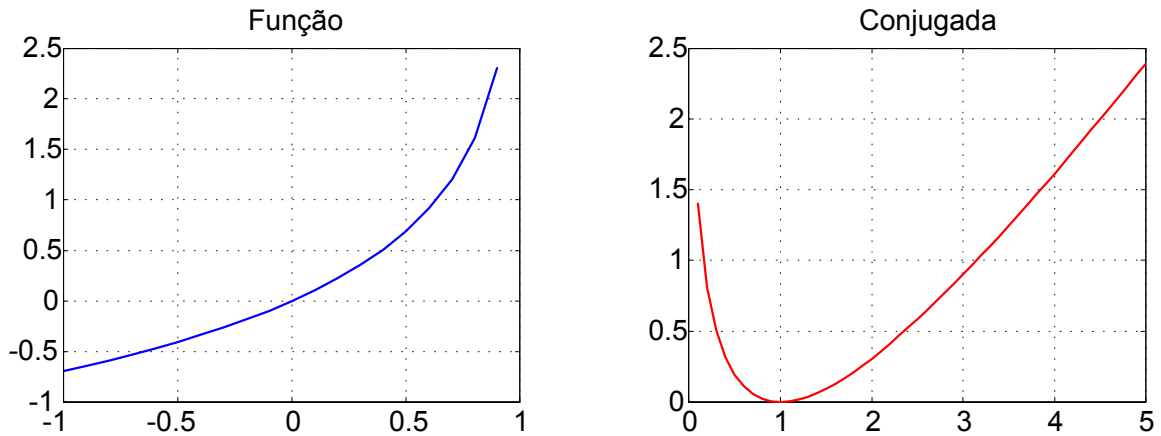


Figura 1.6: Função Barreira Logarítmica e sua Conjugada

A figura(1.6) mostra o gráfico da função f e da sua conjugada f^* . A função conjugada da figura (1.6) é coerciva na fronteira do ortante positivo.

A seguir apresentamos um exemplo em que a função conjugada f^* não é coerciva na fronteira do ortante positivo.

Exemplo 1.9.5 Considere a função $f : \mathcal{R} \rightarrow \mathcal{R} \cup \{+\infty\}$ definida por

$$f(x) = \frac{1}{2}x^2 + x,$$

f tem como conjugada f^* dada por

$$s \in \mathcal{R} \mapsto f^*(s) = \frac{1}{2}(s-1)^2. \quad (1.32)$$

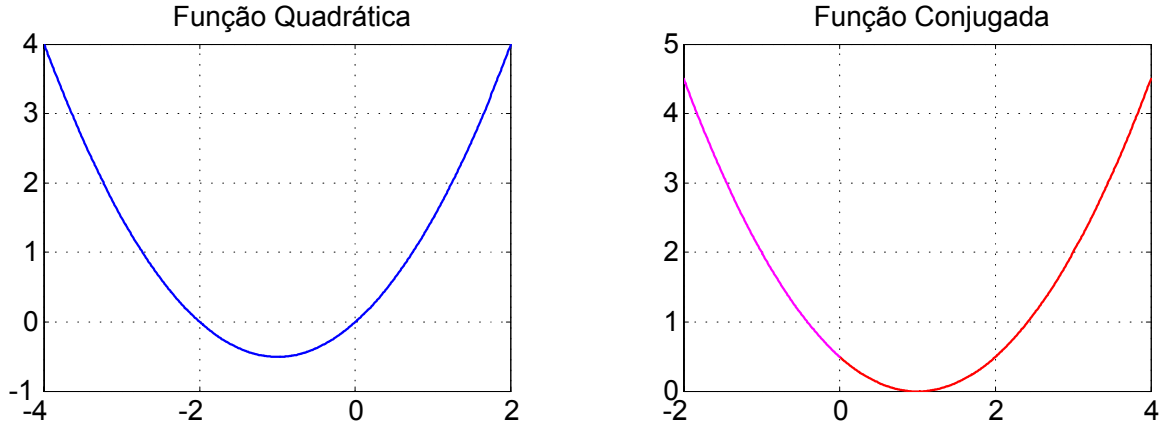


Figura 1.7: Quadrática e sua Conjugada

A figura (1.7) mostra o gráfico da função f e da sua conjugada f^* . A função conjugada da figura (1.7) não é coerciva na fronteira do ortante positivo.

A seguir, apresentaremos algumas propriedades sobre conjugadas que podem ser encontradas em [12] e [21], que serão utilizadas nos próximos capítulos.

Proposição 1.9.6 *Considere as funções $f, f_1, f_2 : \mathcal{R}^n \rightarrow \mathcal{R} \cup \{+\infty\}$ a seguir, convexas e próprias. As seguintes afirmações são verdadeiras:*

- i) A conjugada da função $g(x) = f(x) + \alpha$, é $g^*(s) = f^*(s) - \alpha, \forall \alpha \in \mathcal{R}$;
- ii) A conjugada da função $g(x) = \alpha f(x)$, $\alpha > 0$, é $g^*(s) = \alpha f^*\left(\frac{s}{\alpha}\right)$;
- iii) A conjugada da função $g(x) = f(\alpha x)$, $\alpha \neq 0$, é $g^*(s) = f^*\left(\frac{s}{\alpha}\right)$;
- iv) Se A é um operador linear inversível, então $(f \circ A)^* = f^* \circ (A^{-1})$;
- v) A conjugada da função $g(x) = f(x - x_0)$, é $g^*(s) = f^*(s) + s^T x_0$;
- vi) A conjugada da função $g(x) = f(x) + s_0^T x$, é $g^*(s) = f^*(s - s_0)$;
- vii) Se $f_1 \leq f_2$, então $f_1^* \geq f_2^*$;
- viii) Se $f(x) = \sum_{i=1}^n f_i(x_i)$, com $f_i : \mathcal{R} \rightarrow \mathcal{R} \cup \{+\infty\}$, então $f^*(x) = \sum_{i=1}^n f_i^*(x_i)$.

Exemplo 1.9.7 Considere f a função dada no exemplo (1.9.3) e seja $\alpha > 0$ e $x \in \mathcal{R}^n \rightarrow g(x) = \sum_{i=1}^n f_i(\alpha x_i)$. Então, pelos itens (iii) e (viii) da proposição anterior, juntamente com a conjugada de f dada na relação (1.30), obtemos a conjugada de g

$$s \in \mathcal{R}^n \mapsto g^*(s) = \begin{cases} \sum_{i=1}^n \left(\frac{s_i}{\alpha} \ln \frac{s_i}{\alpha} - \frac{s_i}{\alpha} + 1 \right), & \text{se } s \geq 0 \\ +\infty, & \text{c.c.} \end{cases}.$$

1.10 Quase-distância e Distância de Bregman

Definição 1.10.1 Considere os conjuntos $\mathcal{S} \subset \mathcal{R}^n$ convexo aberto e não vazio e $\overline{\mathcal{S}}$ tais que $\mathcal{S} \subset \overline{\mathcal{S}} \subset \mathcal{R}^n$. Uma função contínua $D: \overline{\mathcal{S}} \times \mathcal{S} \rightarrow \mathcal{R}$ é uma quase-distância se, e somente se, as seguintes condições são satisfeitas

- (a) $D(x, y) \geq 0, \forall (x, y) \in (\overline{\mathcal{S}} \times \mathcal{S})$;
- (b) $D(x, y) = 0 \iff x = y, \forall (x, y) \in (\overline{\mathcal{S}} \times \mathcal{S})$.

A função D , não necessariamente define uma métrica, uma vez que não satisfaz a propriedade de simetria e também não satisfaz a desigualdade triangular. A seguir apresentaremos um exemplo de quase-distância que define uma métrica.

Exemplo 1.10.2 Considere $\mathcal{S} = \mathcal{R}^n$ da definição anterior. A função $D: \mathcal{R}^n \rightarrow \mathcal{R}$ dada por $D(x, y) = \|x - y\|^2$ define uma quase-distância, onde $\|\cdot\|$ é a norma Euclidiana.

Proposição 1.10.3 Considere $D: (\overline{\mathcal{S}} \times \mathcal{S}) \rightarrow \mathcal{R}$ uma quase-distância. Se, para todo $y \in \mathcal{S}$, $D(\cdot, y)$ é convexa, então os conjuntos de nível $\Gamma(y, \alpha) = \{x \in \overline{\mathcal{S}} : D(x, y) \leq \alpha\}$ são limitados, para todo $y \in \overline{\mathcal{S}}$ e todo $\alpha \in \mathcal{R}$.

Prova. Seja $\alpha = 0$. Então, do item (b) da definição (10.1.1), tem-se que o conjunto $\Gamma(x, 0) = \{x \in \overline{\mathcal{S}} : D(x, y) \leq 0\} = \{y\}$, o qual é limitado, para todo $y \in \mathcal{S}$.

Por hipótese $D(\cdot, y)$ é convexa, para todo $y \in \mathcal{R}$. Pelo teorema (1.7.5) tem-se que $\Gamma(\cdot, y)$ é limitado para todo $\alpha \in \mathcal{R}$. ■

Definição 1.10.4 Considere $\mathcal{S} \subset \mathcal{R}^n$ um conjunto aberto e convexo e $\overline{\mathcal{S}}$ tal que $\mathcal{S} \subset \overline{\mathcal{S}} \subset \text{cl}(\mathcal{S})$. Seja $h : \overline{\mathcal{S}} \rightarrow \mathcal{R}$ uma função estritamente convexa e fechada em $\overline{\mathcal{S}}$ e diferenciável em \mathcal{S} . Definimos a distância de Bregman induzida pela função h como

$$(x, y) \in \overline{\mathcal{S}} \times \mathcal{S} \mapsto D_h(x, y) = h(x) - h(y) - \langle \nabla h(y), x - y \rangle \quad (1.33)$$

sendo $\nabla h(\cdot)$ o gradiente da função h , e $\langle \cdot, \cdot \rangle$ o produto interno usual em \mathcal{R}^n .

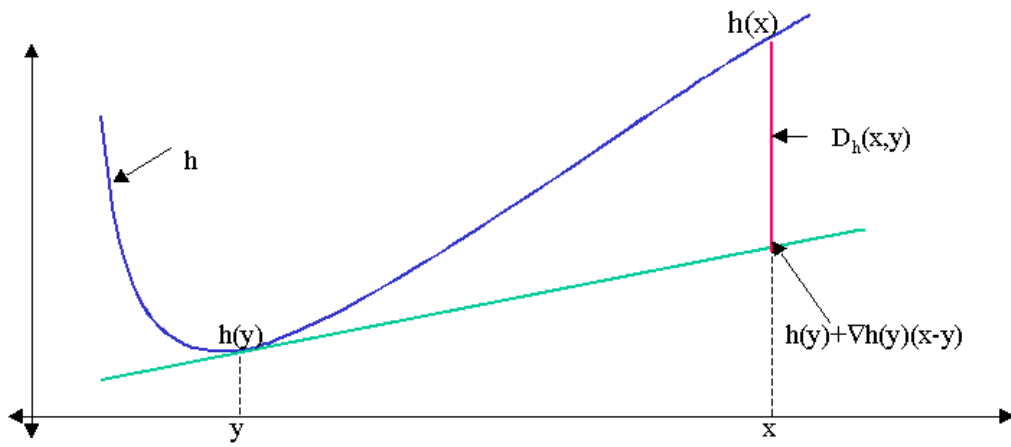


Figura 1.8: Distância de Bregman

Na figura (1.8), pode-se interpretar geometricamente $D_h(x, y)$ como sendo a diferença em x entre h e sua aproximação linear em torno de y .

Na proposição a seguir mostraremos algumas propriedades satisfeitas pela função D_h definida em (1.33), as quais serão usadas posteriormente.

Proposição 1.10.5 Sejam h e D_h dadas como na definição (1.10.4), então

(i) $D_h(x, y) \geq 0$ e $D_h(x, y) = 0$ se, e somente se $x = y$, $\forall x \in \overline{\mathcal{S}}$ e $\forall y \in \mathcal{S}$;

Prova. Como h é estritamente convexa, segue da desigualdade do gradiente, dada na definição(1.7.8), que

$$h(x) - h(y) - \langle \nabla h(y), x - y \rangle > 0$$

e, portanto, da relação (1.33) $D_h(x, y) \geq 0$, $\forall x \in \overline{\mathcal{S}}$ e $\forall y \in \mathcal{S}$. ■

(ii) $D_h(., y)$ é estritamente convexa, $\forall y \in \mathcal{S}$;

Prova. Fixando $y \in \mathcal{R}^n$ e sendo h estritamente convexa, então D_h é escrita como soma de função estritamente convexa com função convexa e, portanto, é estritamente convexa. ■

(iii) Dado $\alpha \in \mathcal{R}$, os conjuntos de nível $\Gamma(x, y) = \{x \in \overline{\mathcal{S}} : D(x, y) \leq \alpha\}$ são limitados, $\forall y \in \mathcal{S}$.

Prova. Dos itens anteriores (i) e (ii) segue que D_h em (1.33) é uma quase-distância, então, pela proposição (1.10.3) fica provado. ■

(iv) Para quaisquer $x \in \overline{\mathcal{S}}$, $y \in \mathcal{S}$ e $z \in \mathcal{S}$, tem-se $D_h(x, y) - D_h(x, z) - D_h(z, y) = \langle \nabla h(y) - \nabla h(z), z - x \rangle$;

Prova. Dada a relação (1.33) tem-se $\forall x \in \overline{\mathcal{S}}$, $\forall y \in \mathcal{S}$ e $\forall z \in \mathcal{S}$:

$$D_h(x, y) = h(x) - h(y) - \langle \nabla h(y), x - y \rangle$$

$$D_h(x, z) = h(x) - h(z) - \langle \nabla h(z), x - z \rangle$$

$$D_h(z, y) = h(z) - h(y) - \langle \nabla h(y), z - y \rangle$$

subtraindo as equações fica

$$D_h(x, y) - D_h(x, z) - D_h(z, y) = -\langle \nabla h(y), x - y \rangle + \langle \nabla h(z), x - z \rangle + \langle \nabla h(y), z - y \rangle$$

agrupando os termos fica

$$D_h(x, y) - D_h(x, z) - D_h(z, y) = \langle \nabla h(y) - \nabla h(z), z - x \rangle$$

$\forall x \in \overline{\mathcal{S}}$, $\forall y \in \mathcal{S}$ e $\forall z \in \mathcal{S}$. ■

(v) Para quaisquer $x, y \in \mathcal{S}$, $\nabla_x D_h(x, y) = \nabla h(x) - \nabla h(y)$.

Prova. Derivando D_h dada na relação (1.33) em relação a variável x tem-se $\nabla_x D_h(x, y) = \nabla h(x) - \nabla h(y)$. ■

Exemplo 1.10.6 Seja $h(x) = \frac{1}{2} \|x\|^2$ neste caso $D_h(x, y) = \frac{1}{2} \|x - y\|^2$. Então os itens (i) à (v) da proposição acima são satisfeitos.

Capítulo 2

Lagrangeano Aumentado

Neste Capítulo é apresentada uma abordagem geral, rápida e objetiva sobre os algoritmos de Lagrangeano Aumentado, Ponto Proximal e Região de Confiança, e as relações existentes entre Ponto Proximal e Lagrangeano Aumentado, e de Ponto Proximal e Região de Confiança. Para isso é necessário um estudo da estrutura de cada um dos métodos utilizados, visando um melhor entendimento dos processos matemáticos envolvidos.

Iniciamos o capítulo apresentando funções de penalidades, definindo-as, e indicando uma metodologia para a construção dessas. As funções de penalidades serão utilizadas no método de Lagrangeano Aumentado.

No método de Ponto Proximal, tem-se que o algoritmo de Ponto Proximal com distância de Bregman, aplicado a um problema definido em um subconjunto do \mathcal{R}^n , está bem definido, no sentido de que dado um ponto inicial é sempre possível determinar o próximo ponto.

Apresentaremos um resultado de equivalência entre o algoritmo de Lagrangeano Aumentado e de Ponto Proximal, o qual garante que a sequência gerada no algoritmo de Lagrangeano Aumentado é a mesma que a gerada por um certo algoritmo de Ponto Proximal.

No final deste capítulo, mostraremos que existe uma certa relação de equivalência entre os métodos de Ponto Proximal e de Região de Confiança, segundo a qual aumentar (ou diminuir) o parâmetro de regularização no método de Ponto Proximal é equivalente a diminuir (ou aumentar) o tamanho da região no método de Região de Confiança.

2.1 Função de Penalidade Coerciva pela Direita

Definimos uma família de funções de penalidades que são coercivas à direita, e posteriormente serão usadas para definir o algoritmo de Lagrangeano Aumentado.

Definição 2.1.1 *Definimos uma família \mathcal{P} de funções de penalidade*

$$y \in \mathcal{R}, \mu \in \mathcal{R}_{++} \longmapsto p(y, \mu) \in \mathcal{R} \cup \{+\infty\} \quad (2.1)$$

com $\text{dom}p(., \mu) = (-\infty, b)$, $b > 0$, possivelmente $b = +\infty$ e para todo $\mu \in \mathcal{R}_{++}$ p . A função satisfaz as seguintes propriedades:

- (P1) $p(0, \mu) = 0$;
- (P2) $p'(0, \mu) = \mu$;
- (P3) $p(., \mu)$ é estritamente convexa e diferenciável em $(-\infty, b)$;
- (P4) $\lim_{y \rightarrow b} p'(y, \mu) = +\infty$;
- (P5) $\lim_{y \rightarrow -\infty} p'(y, \mu) = 0$;

Usamos $p'(y, \mu)$ para denotar a derivada de p em relação à primeira variável.

2.2 Conjugada de uma Penalidade da Família \mathcal{P} e suas Propriedades

Se p é uma função de penalidade da família \mathcal{P} e $\mu \in \mathcal{R}_{++}$ está fixo, então a conjugada de p , segundo a definição (1.9.1), é da forma

$$s \in \mathcal{R} \longmapsto p^*(s, \mu) = \sup\{x^T s - p(x, \mu) : x \in \mathcal{R}\}.$$

Proposição 2.2.1 *Considere $\mu \in \mathcal{R}_{++}$ $p \in \mathcal{P}$. Se p é limitada inferiormente então $\text{dom}p^*(., \mu) = [0, +\infty)$ e se p é ilimitada inferiormente então $\text{dom}p^*(., \mu) = (0, +\infty)$.*

Prova. Ver [17] ■

Da convexidade e diferenciabilidade da função $p(., \mu)$ em $(-\infty, b)$ segue que, $p^*(., \mu)$ é estritamente convexa em $[0, +\infty)$ para o caso limitado inferiormente e $(0, +\infty)$ para o caso ilimitado inferiormente (ver teorema 4.1.3 pág. 81 em [12]).

Da estrita convexidade de $p(., \mu)$ segue, para todo $y \in (-\infty, b)$ e $\mu \in \mathcal{R}_{++}$ que $p^*(., \mu)$ é continuamente diferenciável em $(0, +\infty)$ (ver teorema 4.1.1, pág. 79 em [12]).

Notação 2.2.2 Até o final do capítulo usaremos $\mathcal{S} = \mathcal{R}_{++}^m$ para o caso em que p é limitada inferiormente e $\overline{\mathcal{S}} = \mathcal{R}_+^m$ para o caso em que p é ilimitada inferiormente.

Definição 2.2.3 Uma função $h : \mathcal{R}^n \rightarrow \mathcal{R} \cup \{+\infty\}$ é separável se $h(x) = \sum_{i=1}^m h_i(x_i)$, onde $h_i : \mathcal{R} \rightarrow \mathcal{R} \cup \{+\infty\}$, para todo $i = 1, 2, 3, \dots, m$.

Proposição 2.2.4 Considere p uma penalidade da família \mathcal{P} . Então a função separável D , definida em função da conjugada de p pela fórmula

$$s \in \overline{\mathcal{S}}, \mu \in \mathcal{S} \mapsto D(s, \mu) = \sum_{i=1}^m p^*(s_i, \mu_i), \quad (2.2)$$

é uma quase-distância em $\overline{\mathcal{S}} \times \mathcal{S}$.

Prova. Ver [17] ■

2.3 Construindo Penalidades da Família \mathcal{P}

Na seção anterior vimos como são as funções de penalidades que vamos utilizar na definição do algoritmo de Lagrangeano Aumentado. Em seguida veremos metodologias para construir penalidades da família \mathcal{P} . Basicamente, todas usam uma função auxiliar de uma variável real.

Definição 2.3.1 Considere $\theta : \mathcal{R} \rightarrow \mathcal{R} \cup \{+\infty\}$ uma função satisfazendo as seguintes propriedades:

- i) $\text{intdom } \theta = (-\infty, b)$, $b > 0$, possivelmente $b = +\infty$;
- ii) $\theta(0) = 0$ e $\theta'(0) = 1$;
- iii) θ é estritamente convexa e diferenciável em $(-\infty, b)$;
- iv) $\lim_{y \rightarrow b} \theta'(y) = +\infty$;
- v) $\lim_{y \rightarrow -\infty} \theta'(y) = 0$.

Exemplo 2.3.2 A função exponencial $y \in \mathcal{R} \mapsto \theta(y) = e^y - 1$ satisfaz todas as cinco propriedades acima citadas e este é um caso em que $b = +\infty$, isto é, $\text{intdom } \theta = (-\infty, +\infty)$.

Em nosso trabalho, de agora em diante, trataremos das penalidades da família \mathcal{P} , dividida em 2 tipos, denominando-as de penalidade **tipo 1** e **tipo 2**. Primeiramente trabalharemos com a penalidade **tipo 1**.

Tipo 1: Considere $\theta : \mathcal{R} \rightarrow \mathcal{R} \cup \{+\infty\}$ satisfazendo as propriedades (i) a (v) da definição (2.3.1). Então p dada por

$$(y, \mu) \in \mathcal{R} \times \mathcal{R}_{++} \mapsto p(y, \mu) = \theta(\mu y) \in \mathcal{R} \cup \{+\infty\} \quad (2.3)$$

é uma penalidade da família \mathcal{P} .

A conjugada de p , segundo o item (iii) da proposição (1.9.6), é dada por

$$(s, \mu) \in \mathcal{R} \times \mathcal{R}_{++} \mapsto p^*(s, \mu) = \theta^*\left(\frac{s}{\mu}\right) \in \mathcal{R} \cup \{+\infty\}. \quad (2.4)$$

Exemplo 2.3.3 Considere $\theta : \mathcal{R} \rightarrow \mathcal{R} \cup \{+\infty\}$ definida por

$$\theta(y) = e^y - 1$$

exibida anteriormente no exemplo (2.3.2). Assim p obtida a partir de θ fica

$$y \in \mathcal{R}, \mu \in \mathcal{R}_{++} \mapsto p(y, \mu) = e^{\mu \cdot y} - 1$$

e a conjugada de p , para esse caso, é da forma

$$(s, \mu) \in \mathcal{R} \times \mathcal{R}_{++} \mapsto p^*(s, \mu) = \begin{cases} \left(\frac{s}{\mu}\right) \log\left(\frac{s}{\mu}\right) - \left(\frac{s}{\mu}\right) + 1, & \text{se } s \geq 0 \\ +\infty & , \text{ se } s < 0 \end{cases}.$$

A penalidade **tipo 1** da relação (2.3) pode ser encontrada em [17].

Tipo 2: Considere $\theta : \mathcal{R} \rightarrow \mathcal{R} \cup \{+\infty\}$ satisfazendo as propriedades (i) a (v) da definição (2.3.1). Então p dada por

$$(y, \mu) \in \mathcal{R} \times \mathcal{R}_{++} \mapsto p(y, \mu) = \mu\theta(y) \in \mathcal{R} \cup \{+\infty\} \quad (2.5)$$

é uma penalidade da família \mathcal{P} .

A conjugada de p , segundo o item (ii) da proposição (1.9.6), é dada por

$$(s, \mu) \in \mathcal{R} \times \mathcal{R}_{++} \mapsto p^*(s, \mu) = \mu \theta^* \left(\frac{s}{\mu} \right) \in \mathcal{R} \cup \{+\infty\} \quad (2.6)$$

Exemplo 2.3.4 Considere $\theta : \mathcal{R} \rightarrow \mathcal{R} \cup \{+\infty\}$ definida por $\theta(y) = e^y - 1$. Assim p dada em (2.5) obtida a partir de θ fica

$$y \in \mathcal{R}, \mu \in \mathcal{R}_{++} \mapsto p(y, \mu) = \mu(e^y - 1)$$

a conjugada de p , para esse caso, é da forma

$$(s, \mu) \in \mathcal{R} \times \mathcal{R}_{++} \mapsto p^*(s, \mu) = \begin{cases} s \log \left(\frac{s}{\mu} \right) - s + \mu, & \text{se } s \geq 0 \\ +\infty & , \text{ se } s < 0 \end{cases}.$$

A penalidade **tipo 2** da relação (2.5) é muito utilizada na literatura, e mais informações podem ser obtidas em [1].

Note que a diferença entre as penalidades **tipo 1** e **2** está na posição que o parâmetro $\mu \in \mathcal{R}_{++}$ (multiplicador de Lagrange) ocupa. Nas penalidades **tipo 1**, μ multiplica o argumento da função θ , já nas penalidades do **tipo 2**, μ multiplica a função θ . Mais adiante vamos ver como isso influencia o método.

Apresentaremos em seguida o método de Lagrangeano Aumentado.

2.4 Método de Lagrangeano Aumentado

Nesta seção trataremos sobre a classe de métodos Lagrangeano Aumentado. Lembremos que na sua forma original foram introduzidos para resolver problemas com restrições de igualdade e posteriormente generalizados para problemas com restrições de desigualdade. Neste trabalho, iremos aplicá-los a problemas com restrições de desigualdade.

Considere o seguinte problema

$$\begin{aligned}
 & \text{Minimizar } f(x) \\
 & \text{s.a. } g_i(x) \leq 0, \quad i = 1, \dots, m \\
 & x \in \mathcal{R}^n
 \end{aligned} \tag{P} \tag{2.7}$$

onde as funções f e g_i para $i = 1, 2, \dots, m$ são contínuas e possuem derivadas, pelo menos até primeira ordem, contínuas.

A função Lagrangeana para o problema (2.7) é dada por

$$(x, \mu) \in \mathcal{R}^n \times \mathcal{R}^m \longmapsto \ell(x, \mu) = f(x) + \sum_{i=1}^m \mu_i \cdot g_i(x) \tag{2.8}$$

sendo μ o multiplicador de Lagrange.

A função Lagrangeano Aumentado para o problema (2.7) pode ser escrita da seguinte forma

$$(x, \mu, r) \in \mathcal{R}^n \times \mathcal{R}_+^m \times \mathcal{R}_{++} \longmapsto L(x, \mu, r) = f(x) + r \sum_{i=1}^m p\left(\frac{g_i(x)}{r}, \mu_i\right) \tag{2.9}$$

sendo r o parâmetro de penalização e p uma função de penalidade da família \mathcal{P} .

As condições de Karush-Kuhn-Tucker (K.K.T.) para o problema (2.7), ou sejam, as condições necessárias de primeira ordem para um ponto x^* ser um ponto de mínimo local do problema (2.7), são as seguintes:

$$\begin{aligned}
 \nabla f(x^*) + \sum_{i=1}^m \mu_i^* \cdot \nabla g_i(x^*) &= 0 && \text{(otimalidade)} \\
 \mu_i^* \cdot g_i(x^*) &= 0 && \text{(complementaridade)} \\
 g_i(x^*) &\leq 0 && \text{(restrições)} \\
 \mu_i^* &\geq 0 && \text{(multiplicador de Lagrange)}
 \end{aligned} \tag{2.10}$$

O algoritmo para o método de Lagrangeano Aumentado com restrições de desigualdade é iterativo e começa com $\mu^0 \in \mathcal{R}_+^m, r^0 \in \mathcal{R}_{++}$.

Algoritmo 2.4.1 Dados $\mu^0 \geq 0$ e $r^0 > 0$

faça $k = 0$

repita

encontre

$$x^{k+1} \in \arg \min \left\{ f(x) + r^k \sum_{i=1}^m p \left(\frac{g_i(x)}{r^k}, \mu_i^k \right) \right\}$$

atualize

$$u_i^{k+1} = p' \left(\frac{g_i(x^{k+1})}{r^k}, \mu_i^k \right) \quad i = 1, 2, \dots, m$$

$$k = k + 1$$

A cada iteração o algoritmo de Lagrangeano Aumentado consiste em resolver um subproblema irrestrito, que é a minimização da função Lagrangeano Aumentado dada na relação (2.9), e em seguida atualiza-se os parâmetros.

A sequência (x^k) gerada pelo algoritmo está bem definida (ver afirmação 4.21, pag. 41 em [6]).

Observe que a tarefa pesada do algoritmo é a minimização da função Lagrangeano Aumentado, ao que, em geral, chamamos de subproblema gerado pelo algoritmo. A metodologia do algoritmo é bastante simples. Cada iteração consiste em formar os subproblemas irrestritos, que podem ser resolvidos por qualquer método de minimização irrestrita, e atualizar os parâmetros.

Uma das vantagens de utilizar o método de Lagrangeano Aumentado é que a convergência do método não exige nenhuma restrição quanto ao parâmetro de penalidade, podendo diminuir, aumentar ou manter constante, porém, nos métodos de penalidades clássicos isso não ocorre, exigindo restrições sobre o parâmetro de penalidade para obter convergência. Os métodos de Lagrangeano Aumentado têm por objetivo conciliar os aspectos de mal-condicionamento e evitar perda de estrutura de minimização proveniente do método de penalização [16].

No algoritmo, a cada iteração o ponto x^{k+1} é determinado por

$$x^{k+1} \in \arg \min \left\{ f(x) + r^k \sum_{i=1}^m p \left(\frac{g_i(x)}{r^k}, \mu_i^k \right) \right\}. \quad (2.11)$$

Não conhecemos μ^* antecipadamente, por isso fornecemos inicialmente um valor arbitrário para o vetor μ (normalmente vetor nulo ou vetor de uns). Então esses valores são

atualizados forçando a satisfazer a primeira e a última relação de K.K.T. dadas em (2.10), como mostraremos a seguir.

Derivando, em relação à variável x , a função Lagrangeano Aumentado dada na relação (2.9) tem-se

$$(x, \mu, r) \in \mathcal{R}^n \times \mathcal{R}_+^m \times \mathcal{R}_{++} \longmapsto \nabla_x \ell(x, \mu, r) = \nabla f(x) + r^k \sum_{i=1}^m p' \left(\frac{g_i(x)}{r^k}, \mu_i^k \right) \nabla g_i(x). \quad (2.12)$$

Se escolhermos

$$\mu^{k+1} = p' \left(\frac{g_i(x)}{r^k}, \mu_i^k \right) \quad (2.13)$$

em (2.12) tem-se

$$\nabla f(x) + r^k \underbrace{\sum_{i=1}^m p' \left(\frac{g_i(x)}{r^k}, \mu_i^k \right)}_{u^{k+1}} \nabla g_i(x). \quad (2.14)$$

Observe que (2.14) é a primeira relação de K.K.T. em (2.10). Além disso sendo $p \in \mathcal{P}$ uma penalidade da família \mathcal{P} , é conhecido que $p'(\cdot, \mu^k) \geq 0$, ou seja, $\mu^{k+1} \geq 0$. Assim, os algoritmos de Lagrangeano Aumentado são construídos forçando-se a satisfazer algumas das condições de otimalidade.

Na próxima seção apresentaremos o método de Ponto Proximal.

2.5 Ponto Proximal

Métodos de Ponto Proximal podem ser baseados em operadores monótonos maximais, ou em distâncias generalizadas, por exemplo, distância de Bregman e φ -divergências. Nesta seção, introduzimos o algoritmo de Ponto Proximal com distâncias de Bregman, para programação convexa. Os resultados apresentados aqui serão baseados em Iusem [14]. Salientamos que a distância de Bregman considerada por Iusem é mais restritiva do que aquela que apresentamos na definição (1.10.4), pois ele exige que outras propriedades

sejam satisfeitas na definição de Distância de Bregman.

Na próxima seção, onde será apresentada a relação de equivalência entre Lagrangeano Aumentado e Ponto Proximal, o método de Ponto Proximal será aplicado ao problema dual (a ser definido na próxima seção) com uma quase-distância.

Considere o problema

$$\begin{aligned} & \text{Minimizar } f(x) \\ & \text{s.a. } x \in \overline{\mathcal{S}} \end{aligned} \tag{2.15}$$

sendo $\mathcal{S} \subset \overline{\mathcal{S}} \subset \mathcal{R}^n$ um conjunto aberto e convexo, $\overline{\mathcal{S}}$ o fecho de \mathcal{S} e f convexa e contínua em $\overline{\mathcal{S}}$.

O Método de Ponto Proximal com distância de Bregman consiste em dado

$$x^0 \in \mathcal{S} \tag{2.16}$$

gerar o próximo ponto

$$x^{k+1} \in \arg \min_{x \in \overline{\mathcal{S}}} \{f(x) + r^k D_h(x, x^k)\} \tag{2.17}$$

onde D_h é uma distância de Bregman em \mathcal{S} , dada em [14] e o parâmetro de regularização r^k é positivo para todo $k \in \mathcal{N}$.

Os métodos de Ponto Proximal fazem uma espécie de regularização, ou seja, soma-se à função objetivo um termo positivo, geralmente chamado de núcleo. Quando são aplicados ao problema dual do problema de programação convexa, funcionam como uma espécie de barreira, pois forçam os pontos gerados a ficarem no ortante positivo [14].

Nas últimas décadas, houve um progresso considerável em relação à teoria de métodos de Ponto Proximal baseada em distâncias generalizadas. O grande desafio foi a substituição dos núcleos quadráticos por núcleos não quadráticos. Atualmente, as classes mais utilizadas de métodos não quadráticos são: distâncias de Bregman e φ -divergências.

A seguir apresentaremos o algoritmo de Ponto Proximal ao problema (2.15), com distância de Bregman dada em Iusem [14].

Algoritmo 2.5.1 Dados $x^0 \in \mathcal{S}$, $\bar{r} > 0$ e (r^k) uma sequência em $(0, \bar{r})$

faça $k = 0$

repita

encontre

$$x^{k+1} \in \arg \min_{x \in \mathcal{S}} \{f(x) + r^k D_h(x, x^k)\}$$

$k = k + 1$

Teorema 2.5.2 Se o problema (2.15) tem solução e h é uma função convexa, e coerciva na fronteira de S , então a sequência (x^k) gerada pelo algoritmo de Ponto Proximal converge para a solução x^* do problema (2.15)

Prova. Ver [14] ■

Em nosso trabalho, depois de estabelecer as propriedades de convergência do método de Ponto Proximal, utilizaremos as propriedades que interessam para relacioná-lo com os métodos de Lagrangeano Aumentado e de Região de Confiança. Iusem [14] chama a atenção de que o método de Ponto Proximal deve ser visto como um esquema conceitual antes do que como um algoritmo implementável.

2.6 Relação entre Lagrangeano Aumentado e Ponto Proximal

A relação primal-dual é uma relação envolvendo o algoritmo de Lagrangeano Aumentado, aplicado ao problema primal (2.7), e o algoritmo de Ponto Proximal, aplicado ao problema dual, com uma quase-distância dada pela conjugada da penalidade utilizada no primal. A seguir definiremos o problema dual de (2.7) e apresentaremos o resultado de equivalência entre Lagrangeano Aumentado e Ponto Proximal.

Definição 2.6.1 Associando ao problema (2.7) definimos o problema dual da seguinte forma

$$\begin{aligned} & \text{Minimizar} \quad -d(\mu) \\ & \text{s.a} \quad \mu \geq 0 \end{aligned} \tag{D} \tag{2.18}$$

onde

$$\mu \in \mathcal{R}_+^m \rightarrow d(\mu) = \inf\{\ell(x, \mu) : x \in \mathcal{R}^n\} \quad (2.19)$$

é a função dual Lagrangeana, e ℓ é a função Lagrangeana dada na relação (2.8), ou seja,

$$(x, \mu) \in \mathcal{R}^n \times \mathcal{R}^m \mapsto \ell(x, \mu) = f(x) + \sum_{i=1}^m \mu_i g_i(x).$$

Optamos por $-d(\cdot)$ no problema dual (2.18) para manter o padrão de problema convexo em todo o trabalho.

Em seguida, reescrevemos o algoritmo de Ponto Proximal (2.5.1), agora aplicado ao problema (D) e usando a quase-distância dada na relação (2.2).

Algoritmo 2.6.2 Dados $\mu^0 \in \mathcal{S}$, $\bar{r} > 0$ e (r^k) uma sequência em $(0, \bar{r})$,

faça $k = 0$

repita

encontre

$$\mu^{k+1} \in \arg \min\{-d(\mu) + r^k D(\mu, \mu^k) : \mu \in \overline{\mathcal{S}}\}$$

$$k = k + 1$$

Seja (μ^k) a sequência gerada pelo algoritmo (2.6.2). Assim

$$-d(\mu^{k+1}) + r^k D(\mu^{k+1}, \mu^k) \leq -d(\mu^k) + r^k D(\mu^k, \mu^k).$$

Como D é uma quase-distância e pela definição (1.10.1), $D(\mu^k, \mu^k) = 0$, e além disso, $(-d(\mu^k))$ é uma sequência decrescente, logo

$$-d(\mu^{k+1}) + r^k D(\mu^{k+1}, \mu^k) \leq -d(\mu^k).$$

A questão de dualidade está muito presente no nosso trabalho (às vezes implicitamente). Por um lado, está o método de Lagrangeano Aumentado, por outro lado o método de Ponto Proximal. O próximo teorema relaciona os métodos de Lagrangeano Aumentado e de Ponto Proximal, e para isso, consideraremos as seguintes hipóteses:

- Hipótese **H1**: O conjunto solução Ω^* do Problema (2.7) é não vazio e compacto;
- Hipótese **H2**: Existe $\bar{x} \in \text{dom} f_0$ tal que $g_i(\bar{x}) < 0$, para todo $i = 1, 2, \dots, m$.
(conhecida na literatura como condição de qualificação de Slater).

Teorema 2.6.3 *Considere o problema (2.7) satisfazendo as hipóteses **H1** e **H2**, (x^k, μ^k) seqüências geradas pelo algoritmo de Lagrangeano Aumentado (2.4.1), aplicado ao problema (2.7) com penalidade $p \in \mathcal{P}$, e $(\bar{\mu}^k)$ uma seqüência gerada pelo algoritmo de Ponto Proximal (2.6.2), aplicado ao problema (2.18) com quase-distância dada pela conjugada de p , determinada na relação (2.2). Se $\mu^0 = \bar{\mu}^0$, então $\mu^k = \bar{\mu}^k$, para todo $k \in \mathcal{N}$.*

Prova. ver [6] ■

O teorema (2.6.3) nos diz que a seqüência de multiplicadores gerada no algoritmo de Lagrangeano Aumentado é a mesma que a gerada pelo algoritmo de Ponto Proximal, quando este é aplicado ao dual do problema (2.7) e com conjugada da penalidade utilizada no algoritmo de Lagrangeano Aumentado.

No próximo teorema, é mostrado que o algoritmo de Lagrangeano Aumentado (2.4.1), com penalidade p **tipo 1**, dada na relação (2.3), gera seqüências (x^k) e (μ^k) que satisfazem a condição de complementaridade dada na relação (2.10).

Teorema 2.6.4 *Considere (x^k) e (μ^k) seqüências geradas pelo algoritmo de Lagrangeano Aumentado (2.4.1) com penalidade **tipo 1**, dada na relação (2.3), e (r^k) uma seqüência em (r_1, r_2) tais que $0 < r_1 < r_2$. Nestas condições, tem-se*

$$\lim_{k \rightarrow \infty} \frac{u_i^k \cdot g_i(x^{k+1})}{r^k} = 0, \quad \forall i = 1, 2, \dots, m.$$

Prova. Ver [17] ■

2.7 Relação entre Ponto Proximal e Região de Confiança

O método de Região de Confiança aplicado ao problema irrestrito é um procedimento iterativo que a cada iteração consiste em transformar o problema original num

problema mais simples. Dessa forma, em cada iteração, constrói-se um modelo local da função, que é minimizado em uma certa região, a qual é chamada de Região de Confiança. Nesta seção, consideraremos o caso em que o modelo é exato, no sentido que minimizamos a própria função e não o modelo, isso porque nosso interesse, neste momento, é um estudo teórico relativo a esse método. Mostraremos que “a menos de uma constante” os subproblemas resolvidos são “equivalentes” (num sentido a ser definido a seguir) aos subproblemas resolvidos pelo método de Ponto Proximal.

O problema a ser tratado nesta seção tem a forma

$$\begin{aligned} & \text{Minimizar } f(x) \\ & x \in \mathcal{R}^n \end{aligned} \tag{2.20}$$

onde $f : \mathcal{R}^n \rightarrow \mathcal{R}$ é uma função convexa.

Uma iteração k do método de Região de Confiança, aplicado ao problema (2.20), consiste em, partindo de x^k e $\Delta_k > 0$, determinar uma solução do subproblema

$$x^{k+1} \in \arg \min_{x \in \mathcal{R}^n} \{f(x) : D(x, x^k) \leq \Delta_k\} \tag{2.21}$$

sendo $D(., x^k)$ uma quase-distância definida em (1.10.1), por exemplo, $x \in \mathcal{R}^n \mapsto D(x, x^k) = \|x - x^k\|^2$, onde $\|\cdot\|$ é a norma Euclidiana, muito utilizada nos métodos de Região de Confiança e Pontos Proximais tradicionais (núcleos quadráticos).

Da mesma forma, uma iteração k do método de Ponto Proximal, aplicado ao problema (2.20), consiste em, partindo de $x^k > 0$, $r^k > 0$ e uma quase-distância $D(., x^k)$ dada, determinar uma solução do subproblema

$$x^{k+1} \in \arg \min_{x \in \mathcal{R}^n} \{f(x) + r^k D(x, x^k)\}. \tag{2.22}$$

Note que, no algoritmo de Ponto Proximal (2.5.1) $x \in \overline{\mathcal{S}}$. Na relação (2.22) $x \in \mathcal{R}^n$ pois se trata do problema irrestrito (2.20).

A próxima proposição é referente à equivalência existente entre (2.21) e (2.22), no sentido que sendo x^* uma solução do problema (2.21), existe $r^k > 0$ tal que x^* é uma solução do problema (2.22). Da mesma forma, se x^* uma solução do problema (2.22)

existe $\Delta_k \geq 0$, tal que, x^* é uma solução do problema (2.21).

Proposição 2.7.1 *Os problemas (2.21) e (2.22) são equivalentes, no sentido que, sendo x^* uma solução do problema (2.21), existe $r^k \geq 0$ tal que x^* é uma solução do problema (2.22). Da mesma forma, se x^* uma solução do problema (2.22), existe $\Delta_k \geq 0$ tal que x^* uma solução do problema (2.21).*

Prova. *Seja D uma quase-distância, então $D(x^k, x^k) = 0 \leq \Delta_k$. Dado que x^* é solução do (2.21), existe $\mu \geq 0$ tal que x^* é uma solução do seguinte problema*

$$\text{Minimizar} \{ \ell(x, \mu) : x \in \mathcal{R}^n \} \quad (2.23)$$

onde

$$x \in \mathcal{R}^n, \mu \in \mathcal{R}_+^m \mapsto \ell(x, \mu) = f(x) + \mu(D(x, x^k) - \Delta_k) \quad (2.24)$$

é a função Lagrangeana associada ao problema (2.21).

Portanto, a primeira parte está demonstrada, pois uma solução do problema (2.23) também é solução do problema (2.22). A seguir faremos a segunda parte da demonstração.

Sejam $x \in \mathcal{R}^n$, $r^k \geq 0$ e $D(., x^k)$ uma quase-distância. Se x^* é uma solução do problema (2.22), então também o será do seguinte problema

$$\underset{x \in \mathcal{R}^n}{\text{minimizar}} \{ f(x) + r^k D(x, x^k) - r^k \Delta_k \} \quad (2.25)$$

sendo $\Delta_k \in \mathcal{R}_+$.

Considerando

$$r \in \mathcal{R}_+ \mapsto d(r) = \inf \{ \ell(x, r) : x \in \mathcal{R}^n \} \quad (2.26)$$

sendo ℓ a função Lagrangeana (2.24) com r no lugar de μ .

Como x^* é solução do problema (2.25), tem-se que

$$d(r^k) = f(x^*) + r^k (D(x^*, x^k) - \Delta_k). \quad (2.27)$$

Fazendo $\Delta_k = D(x^*, x^k)$ na relação (2.27), obtém-se $d(r^k) = f(x^*)$, pelo teorema forte de dualidade, x^* é uma solução do problema (2.21). ■

No próximo lema, vamos mostrar que aumentando-se (diminuindo-se) o parâmetro r no problema (2.22), então diminui-se (aumenta-se) a Região de Confiança no problema (2.21).

Lema 2.7.2 *Considere x_1 e x_2 soluções do problema (2.22), r_1 e r_2 as respectivas constantes associadas a essas soluções, tal que $r_1 > r_2 > 0$, então*

$$D(x_1, x^k) \leq D(x_2, x^k) \quad (2.28)$$

Prova. *Considere $x^k \in \mathcal{R}^n$ e x_1 e x_2 soluções do problema (2.22), tal que r_1 e r_2 são as respectivas constantes associadas a essas soluções e $r_1 > r_2 > 0$. Como x_1 é solução do problema (2.22) então*

$$g(x_1) + r_1 D(x_1, x^k) \leq g(x_2) + r_2 D(x_2, x^k). \quad (2.29)$$

Da mesma forma, se x_2 é solução do problema (2.22) então

$$g(x_2) + r_2 D(x_2, x^k) \leq g(x_1) + r_1 D(x_1, x^k). \quad (2.30)$$

Somando as desigualdades (2.29) e (2.30) e agrupando os termos tem-se

$$(r_1 - r_2) D(x_1, x^k) \leq (r_1 - r_2) D(x_2, x^k). \quad (2.31)$$

Como $r_1 > r_2 > 0$, segue que

$$D(x_1, x^k) \leq D(x_2, x^k).$$

■

Esse resultado que relaciona Ponto Proximal e Região de Confiança será importante para os próximos capítulos, principalmente para o capítulo onde estudaremos a implementação do algoritmo de Lagrangeano Aumentado.

Capítulo 3

O Caso Linear

Começamos este capítulo definindo o problema linear e seu respectivo dual. Acreditamos que se um algoritmo apresentar boas propriedades no caso linear, terá alguma chance de ter boas propriedades no caso não-linear. Agora, se um algoritmo não possui um bom desempenho para problemas de programação linear, também não deverá possuir para problemas de programação não-linear.

Em seguida apresentaremos os algoritmos de Lagrangeano Aumentado aplicados ao problema primal linear e de Ponto Proximal aplicado ao problema dual linear com função de penalidade quadrática.

O resultado principal do capítulo é que as direções duais geradas pelo algoritmo de Lagrangeano Aumentado, aplicado ao problema linear e penalidade **tipo 1** quadrática são equivalentes às geradas pelo algoritmo Afim Escala aplicado ao dual do problema linear. Esse resultado também é muito bom, porque controlando adequadamente o parâmetro de penalidade no algoritmo de Lagrangeano Aumentado, os multiplicadores gerados por este são os mesmos que os gerados pelo algoritmo Afim Escala. Porém esse resultado não se aplica para as penalidades **tipo 2**.

3.1 Programação Linear

A programação ou otimização linear está ligada à solução de um tipo muito especial de problema, que consiste na minimização ou maximização de uma função linear podendo ter restrições lineares de igualdade e/ou desigualdade.

Pode-se definir o problema geral de Programação linear da seguinte forma: dado um conjunto de "m" desigualdades ou equações lineares em "n" variáveis, queremos determinar os valores não negativos dessas variáveis que satisfaçam as restrições e minimizam alguma função linear.

Uma forma padrão do problema de programação linear com 'm' restrições e 'n' variáveis com notação matricial pode ser representada por

$$\begin{aligned}
 & \text{Minimizar } f(x) = c^T x \\
 & \text{s.a} \\
 & \quad Ax = b \\
 & \quad x \geq 0
 \end{aligned} \tag{3.1}$$

onde $A \in \mathcal{R}^{m \times n}$, $c \in \mathcal{R}^n$, $b \in \mathcal{R}^m$, isto é, A é uma matriz $m \times n$, com $m \geq n$, c um vetor $n \times 1$ e b um vetor $m \times 1$. A notação $x \geq 0$ quer dizer: $x_i \geq 0$, para todo $i = 1, \dots, n$. A função linear $f : \mathcal{R}^n \rightarrow \mathcal{R}$ é chamada função objetivo. Uma solução que satisfaça as restrições $Ax = b$, $x \geq 0$ é dita solução viável. A cada modelo de programação linear, contendo coeficientes da matriz A , os coeficientes dos vetores b e c , corresponde a um outro modelo, denominado dual, formado por esses mesmos coeficientes, porém disposto de maneira diferente. Ao modelo original dá-se o nome de primal. De (3.1) tem-se o dual escrito da seguinte forma:

$$\begin{aligned}
 & \text{Minimizar } f(x) = -b^T y \\
 & \text{s.a} \\
 & \quad Ay \leq c \\
 & \quad y \text{ irrestrito de sinal.}
 \end{aligned} \tag{3.2}$$

A maior parte do desenvolvimento teórico de algoritmos para programação linear é feita sobre o enunciado com restrição de igualdade [9], também chamado de formato primal ou de formato padrão. Em geral, o enunciado com restrição de desigualdade é chamado de formato dual. Um problema é enunciado em formato primal, seu dual estará naturalmente em formato dual, e vice-versa.

Para o desenvolvimento deste trabalho, vamos utilizar o formato dual para o

problema primal, sendo escolhido desse modo com o objetivo de manter a notação que vimos fazendo até aqui, a mesma do caso não linear. Assim, definimos o problema

$$\begin{aligned}
 & \text{Minimizar} && -b^T x \\
 & \text{sa} && \\
 & && Ax \leq c \\
 & && x \in \mathcal{R}^n
 \end{aligned} \tag{PL} \tag{3.3}$$

sendo $A \in \mathcal{R}^{m \times n}$, $b \in \mathcal{R}^n$, $c \in \mathcal{R}^m$ e $m \geq n$.

O sinal negativo na função objetivo é simplesmente uma convenção, para facilitar o desenvolvimento que faremos, lembrando que, se x^* resolve o problema de *maximizar* $f(x)$, também resolverá o de *minimizar* $-f(x)$.

Proposição 3.1.1 *A função dual Lagrangeana associada ao problema (PL) tem a seguinte forma*

$$\mu \in \mathcal{R}_+^m \mapsto d(\mu) = \begin{cases} -c^T \mu, & \text{se } A^T \mu = b \\ -\infty & , \text{ caso contrário} \end{cases} \tag{3.4}$$

Prova. *A função dual Lagrangeana, já definida anteriormente na relação (2.19), para o caso do problema linear, tem a forma*

$$\begin{aligned}
 \mu \in \mathcal{R}_+^m \mapsto d(\mu) &= \inf\{\ell(x, \mu) : x \in \mathcal{R}^n\} = \\
 &= \inf\{-b^T x + \mu^T (Ax - c) : x \in \mathcal{R}^n\} = \\
 &= \inf\{(-b + A^T \mu)^T x - c^T \mu : x \in \mathcal{R}^n\}
 \end{aligned}$$

como $(-b + A^T \mu)^T x - c^T \mu$ é linear em x . Se $-b + A^T \mu \neq 0$ o ínfimo é $-\infty$. No caso contrário, tem-se que $b = A^T \mu$. Assim o ínfimo é $-c^T \mu$ ■

A seguir, definiremos o dual associado ao problema (PL) dado na relação (3.3)

$$\begin{aligned}
 & \text{Minimizar} && c^T \mu \\
 & \text{sa} && A^T \mu = b \\
 & && \mu \geq 0
 \end{aligned} \tag{DL} \tag{3.5}$$

3.2 A Penalidade Quadrática

Nesta seção definiremos a função θ que será utilizada na definição das penalidades quadráticas **tipo 1** e **2**. Dessa forma, definimos

$$y \in \mathcal{R} \mapsto \theta(y) = \frac{1}{2}y^2 + y. \quad (3.6)$$

A conjugada de θ dada na relação (3.6) é da forma

$$s \in \mathcal{R} \mapsto \theta^*(s) = \frac{1}{2}(s - 1)^2. \quad (3.7)$$

Considere a penalidade p **tipo 1** dada em relação (2.3), isto é,

$$(y, \mu) \in \mathcal{R} \times \mathcal{R}_{++} \mapsto p(y, \mu) = \theta(\mu y) \in \mathcal{R} \quad (3.8)$$

e a penalidade p **tipo 2** dada em relação (2.5), isto é,

$$(y, \mu) \in \mathcal{R} \times \mathcal{R}_{++} \mapsto p(y, \mu) = \mu\theta(y) \in \mathcal{R}. \quad (3.9)$$

Para o caso linear, como foi feito anteriormente no caso não linear, aplicando o algoritmo de Ponto Proximal ao problema dual (DL), usando a quase-distância dada pela conjugada da penalidade p **tipo 1** ou **2** e θ^* dada na relação (3.7), os multiplicadores gerados podem ser negativos, o mesmo ocorrendo no primal quando o algoritmo de Lagrangeano Aumentado é aplicado ao problema (PL) (3.3) e θ quadrática dada na relação (3.6).

Temos que θ dada na relação (3.6) não é crescente em todo seu domínio, isto é, no intervalo de $(-\infty, -1)$ é decrescente e no intervalo de $(-1, \infty)$ é crescente. Por não ser crescente em todo o seu domínio gera derivadas direcionais negativas. Com θ dada na relação (3.6) as funções de penalidade construídas a partir dessa não serão da família \mathcal{P} . Observa-se que a propriedade de limitação por baixo da derivada, dada no item (P5) da definição (2.1.1), não é satisfeita aqui, pois $\lim_{y \rightarrow -\infty} \theta'(y) = -\infty$, que deveria ser $\lim_{y \rightarrow -\infty} \theta'(y) = 0$. Podemos notar que θ^* dada na relação (3.7) não é coerciva no ortante não negativo (ver exemplo (1.9.5)). Então os multiplicadores gerados podem ser negativos.

3.3 Ponto Proximal Aplicado ao Problema Dual, com Quase-distância Quadrática

Apresentaremos nesta seção o algoritmo de Ponto Proximal aplicado ao problema (DL) dado na relação (3.5), utilizando uma quase-distância quadrática. Lembrando que, no primal, a penalidade quadrática é formada por uma penalidade p , **tipo 1** ou **2**, e função θ quadrática, dada na relação (3.6). No dual a quase-distância é formada pela conjugada da função θ^* dada na relação (3.7).

As quase-distâncias para as penalidades **tipo 1** e **2** com θ^* dada na relação (3.7), são, respectivamente,

$$(s, \mu) \in \mathcal{R}_+^m \times \mathcal{R}_{++}^m \mapsto D(s, \mu) = \sum_{i=1}^m \frac{1}{2} \left(\frac{s_i}{\mu_i} - 1 \right)^2 \quad (3.10)$$

e

$$(s, \mu) \in \mathcal{R}_+^m \times \mathcal{R}_{++}^m \mapsto D(s, \mu) = \sum_{i=1}^m \frac{\mu_i}{2} \left(\frac{s_i}{\mu_i} - 1 \right)^2. \quad (3.11)$$

A seguir, apresentaremos o algoritmo de Ponto Proximal, aplicado ao problema (DL) com quase-distâncias quadráticas dadas nas relações (3.10) e (3.11).

Algoritmo 3.3.1 Dados $\mu^0 > 0$ e $r^0 > 0$

faça $k = 0$

repita

encontre

$$\mu^+ \in \arg \min_{\mu \geq 0} \{-d(\mu) + r^k D(\mu, \mu^k)\}$$

se $\mu^+ \geq 0$

$$\mu^{k+1} = \mu^+$$

$$r^{k+1} < r^k$$

senão

$$r^{k+1} > r^k$$

$$k = k + 1$$

A metodologia do algoritmo (3.3.1) consiste em a cada iteração resolver o sub-problema irrestrito gerado pela função dual Lagrangeana. Se $\mu^+ \geq 0$, atualiza-se o multiplicador corrente para a próxima iteração e reduz-se o parâmetro de regularização r . Caso contrário, isto é, se $\mu_i^+ < 0$ para algum $i \in \{1, 2, \dots, m\}$, mantém-se o ponto e aumenta-se o valor do parâmetro de regularização r .

Como D é uma quase-distância quadrática, μ^+ gerado na solução do subproblema interno

$$\arg \min_{\mu \geq 0} \{-d(\mu) + r^k D(\mu, \mu^k)\}, \quad (3.12)$$

pode ser negativo, conforme a figura (3.1) a seguir.

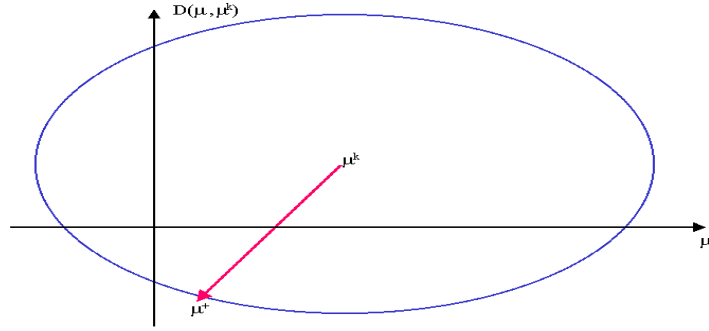


Figura 3.1: Passo do Método de Ponto Proximal

Para contornar o problema de μ^+ não ser positivo, recorreremos à seção 2.7, onde mostramos na proposição (2.7.1) que o problema

$$\arg \min_{\mu \geq 0} \{-d(\mu) + r^k D(\mu, \mu^k)\}, \quad (3.13)$$

exibe uma certa relação de equivalência com subproblema de Região de Confiança

$$\arg \min_{\mu \geq 0} \{-d(\mu) : D(\mu, \mu^k) \leq \Delta_k\}. \quad (3.14)$$

Sabe-se pela proposição (2.7.1) que dados $\mu^k > 0$, $r^k > 0$ e $\bar{\mu}$ uma solução do problema (3.13), existe $\Delta_k > 0$ tal que $\bar{\mu}$ é solução do problema (3.14). No lema (2.7.2) tem-se que aumentando o valor de r^k no problema (3.13) diminui-se a Região de Confiança

do problema (3.14).

Dessa forma, se algumas das componentes de μ^+ no algoritmo de Ponto Proximal (3.3.1) são negativas, como mostra a figura (3.1), aumenta-se o valor do parâmetro r e diminui-se a Região de Confiança, desta maneira μ^+ é forçado a ficar no ortante positivo, conforme a figura (3.2) a seguir. Esse procedimento é repetido até alcançar $\mu^k \geq 0$.

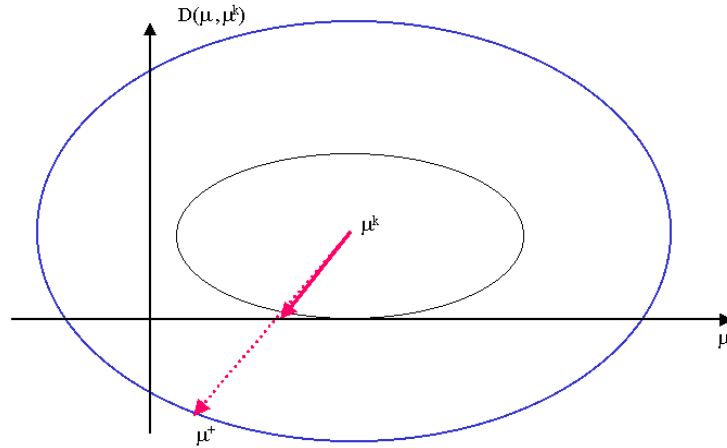


Figura 3.2: Passo do Método de Ponto Proximal Forçando μ Ficar não Negativo

3.4 Lagrangeano Aumentado para o Problema Linear com Penalidade Quadrática

Apresentaremos nesta seção o algoritmo de Lagrangeano Aumentado, aplicado ao problema linear (PL), com penalidade p **tipo 1** ou **tipo 2** dadas, respectivamente, nas relações (3.8) e (3.9) e θ a função quadrática dada na relação (3.6).

No problema (PL) dado na relação (3.3) a função objetivo é

$$\text{Minimizar } f(x) = -b^T x$$

e as restrições são da forma

$$g(x) = Ax - c \leq 0.$$

Assim, pode-se construir a função Lagrangeana

$$(x, \mu) \in \mathcal{R}^n \times \mathcal{R}^m \longmapsto \ell(x, \mu) = -b^T x + \sum_{i=1}^m \mu_i (A_i(x) - c_i), \quad (3.15)$$

em que A_i denota a linha i da matriz A , para $i = 1, \dots, m$.

A função Lagrangeano Aumentado para o problema (3.3) pode ser escrita da seguinte forma

$$(x, \mu, r) \in \mathcal{R}^n \times \mathcal{R}_+^m \times \mathcal{R}_{++} \longmapsto L(x, \mu, r) = f(x) + r^k \sum_{i=1}^m \mu_i p\left(\frac{A_i(x) - c_i}{r^k}, \mu_i\right). \quad (3.16)$$

No algoritmo, a cada iteração, atualiza-se o próximo ponto como sendo

$$x^{k+1} \in \arg \min \left\{ f(x) + r^k \sum_{i=1}^m p\left(\frac{A_i(x) - c_i}{r^k}, \mu_i^k\right) \right\}. \quad (3.17)$$

A seguir escreveremos o algoritmo de Lagrangeano Aumentado e em seguida explicaremos os principais passos desse.

Algoritmo 3.4.1 Dados $\mu^0 \geq e$ e $r^0 > 0$

$k = 0$

repita

encontre

$$x^+ \in \arg \min_{x \in \mathcal{R}^n} \left\{ -b^T x + \sum_{i=1}^m r^k p\left(\frac{A_i(x) - c_i}{r^k}, \mu_i^k\right) \right\}$$

calcule

$$\mu_i^+ = p'\left(\frac{A_i(x^+) - c_i}{r^k}, \mu_i^k\right), \quad i = 1, 2, \dots, m$$

se $\mu^+ \geq 0$

$$x^{k+1} = x^+$$

$$\mu^{k+1} = \mu^+$$

$$r^{k+1} < r^k$$

senão

$$r^{k+1} > r^k$$

$$k = k + 1$$

A metodologia do algoritmo (3.4.1) é bastante simples. Cada iteração consiste em formar subproblemas irrestritos, que podem ser resolvidos por qualquer método de minimização irrestrita, e atualizar os parâmetros. Se $\mu^+ \geq 0$, atualiza-se o ponto corrente, o multiplicador de Lagrange para a próxima iteração e reduz-se o parâmetro de penalidade r . Caso contrário, isto é, $\mu^+ \not\geq 0$, mantém-se o ponto e aumenta-se o valor do parâmetro de penalidade r .

3.5 Afim Escala

O algoritmo Afim Escala foi desenvolvido inicialmente para problemas de programação linear. No caso Afim Escala primal o algoritmo parte de um ponto interior viável, realiza uma mudança de escala com intenção de centralizar o iterando corrente para tornar uma direção de descida conveniente $d \in \mathcal{R}^n$. Essa direção é calculada no espaço afim, minimizando uma função linear em uma bola. Segue-se uma nova mudança de escala para retornar ao espaço original. Isso corresponde a minimizar uma função linear em um elipsóide. O algoritmo Afim Escala que apresentaremos é baseado em [9] e [10], e será aplicado ao problema (DL) dado na relação (3.5).

Algoritmo 3.5.1 *Dados $\mu^0 > 0$ e $\gamma \in (0, 1)$*

faça $k = 0$

repita

encontre

$$\mu^+ \in \arg \min_{\mu \in \mathbb{R}_+^m} \left\{ c^T \mu : \left\| \frac{\mu}{\mu^k} - e \right\| \leq \Delta, A^T \mu = b \right\}, \quad e = (1, 1, \dots, 1)^T$$

calcule

$$h = \mu^+ - \mu^k$$

$$\bar{\lambda} = \max\{\lambda \in \mathcal{R} : \mu^k + \lambda h \geq 0\}$$

$$\mu^{k+1} = \mu^k + \gamma \bar{\lambda} h$$

$$k = k + 1$$

O algoritmo resolve uma seqüência de problemas fáceis, segundo [9], cujas soluções devem convergir para uma solução ótima do problema. Cada iteração inicia com um ponto viável μ^k e minimiza uma aproximação da função objetivo (no caso linear a própria função)

em uma Região de Confiança, gerando uma direção h , seguido de uma busca unidirecional ao longo de h , a partir de μ^k , para determinar o tamanho do passo.

No algoritmo (3.5.1) o maior trabalho é o do cálculo do novo ponto μ^k , que consiste em minimizar, em um hiperplano, uma função linear num elipsóide. Em [9] (Seções 3.3.3 e 3.3.5, pág. 32) o autor fornece uma fórmula explícita para o cálculo desse.

3.6 Equivalência entre Afim Escala e Lagrangeano Aumentado com Penalidade Quadrática

Nesta seção provaremos o resultado mais importante do capítulo, o que relaciona os métodos Afim Escala e Lagrangeano Aumentado, no caso em que a penalidade é do **tipo 1** dada na relação (3.8) e a função θ quadrática dada na relação (3.6). Esse resultado já foi provado em [17], aqui reescreveremos a demonstração readaptando ao nosso trabalho.

Teorema 3.6.1 *O algoritmo de Lagrangeano Aumentado no caso linear (3.4.1), aplicado ao problema (PL) com penalidade p do **tipo 1**, dada na relação (3.8), e θ quadrática dada na relação (3.6), gera direções duais colineares às geradas pelo algoritmo Afim Escala (3.5.1) aplicado ao problema (DL).*

Prova. *Considere $\mu^k > 0$ e $r^k > 0$ e a k -ésima iteração do algoritmo de Ponto Proximal (3.3.1), aplicado ao problema (DL), com quase-distância*

$$\mu \in \mathcal{R}_+^m, \mu^k \in \mathcal{R}_{++}^m \mapsto D(\mu, \mu^k) = \sum_{i=1}^m p^*(\mu_i, \mu_i^k), \quad (3.18)$$

onde p^* é a conjugada de p **tipo 1** dada na relação (3.8), isto é,

$$(y, \mu) \in \mathcal{R} \times \mathcal{R}_{++} \mapsto p(y, \mu) = \theta(\mu y) \in \mathcal{R}.$$

Assim p^* pode ser escrita da seguinte forma:

$$s \in \mathcal{R}_+^m, \mu \in \mathcal{R}_{++}^m \mapsto p^*(s, \mu) = \theta^* \left(\frac{s}{\mu} \right). \quad (3.19)$$

Aplicando θ^* dada na relação (3.7) em (3.19), tem-se

$$s \in \mathcal{R}_+^m, \mu \in \mathcal{R}_{++}^m \mapsto p^*(s, \mu) = \frac{1}{2} \left(\frac{s}{\mu} - 1 \right)^2 \quad (3.20)$$

substituindo (3.20) em (3.18), tem-se

$$\mu \in \mathcal{R}_+^m, \mu^k \in \mathcal{R}_{++}^m \mapsto D(\mu, \mu^k) = \sum_{i=1}^m \frac{1}{2} \left(\frac{\mu_i}{\mu_i^k} - 1 \right)^2.$$

Assim, o ponto μ^{k+1} no algoritmo de Ponto Proximal (3.3.1) é determinado por

$$\mu^{k+1} \in \arg \min \left\{ -d(\mu) + \frac{r^k}{2} \sum_{i=1}^m \left(\frac{\mu_i}{\mu_i^k} - 1 \right)^2 \right\} \quad (3.21)$$

sendo d a função dual Lagrangeana dada na relação (3.4). Reescrevendo a relação (3.21), em termo da norma euclidiana, tem-se

$$\mu^{k+1} \in \arg \min_{\mu \geq 0} \left\{ -d(\mu) + \frac{r^k}{2} \left\| \frac{\mu}{\mu^k} - e \right\|^2 \right\} \quad (3.22)$$

onde o vetor unitário $e = (1, \dots, 1)^T \in \mathcal{R}^m$.

No capítulo 2, mostramos, na proposição (2.7.1), que existe uma equivalência entre Região de Confiança e Ponto Proximal. Deste modo, existe $\Delta_k > 0$ tal que o problema (3.22) é equivalente ao problema de Região de Confiança

$$\mu^{k+1} \in \arg \min_{\mu \geq 0} \left\{ -d(\mu) : \underbrace{\left\| \frac{\mu}{\mu^k} - e \right\|}_{D(u, u^k)} \leq \Delta_k \right\}. \quad (3.23)$$

Usando a função dada pela relação (3.4), no problema (3.23) tem-se

$$\mu^{k+1} \in \arg \min_{\mu \geq 0} \left\{ c^T \mu : \left\| \frac{\mu}{\mu^k} - e \right\| \leq \Delta_k \text{ e } A^T \mu = b \right\}. \quad (3.24)$$

Assim, fica provada a equivalência entre Afim Escala e Ponto Proximal pois

o problema (3.24) é exatamente o mesmo resolvido na k -ésima iteração do algoritmo Afim Escala (3.5.1) aplicado ao problema (DL). Logo, existe $\Delta_k > 0$ tal que μ^{k+1} determinado pelo algoritmo Afim Escala coincide com μ^{k+1} determinado pelo algoritmo de Ponto Proximal. Pelo teorema de equivalência entre Ponto Proximal e Lagrangeano Aumentado (2.6.3) μ^{k+1} também é determinado pelo algoritmo de Lagrangeano Aumentado aplicado ao problema (PL) com penalidade p **tipo 1** e θ quadrática dada na relação (3.6). ■

É importante ressaltar que esse resultado só vale para penalidade do **tipo 1**. Para penalidade **tipo 2**, o resultado não é válido, pois a conjugada é da forma

$$\mu \in \mathcal{R}_+^m, \mu^k \in \mathcal{R}_{++}^m \mapsto D(\mu, \mu^k) = \sum_{i=1}^m \mu_i^k \theta^* \left(\frac{\mu_i}{\mu_i^k} \right).$$

Sendo θ^* quadrática dada na relação (3.7), repetindo a prova do teorema (3.6.1), ao invés da relação (3.21) teremos

$$\mu^{k+1} \in \arg \min_{\mu \geq 0} \left\{ -d(\mu) + \frac{r^k}{2} \sum_{i=1}^m \mu_i^k \left(\frac{\mu_i}{\mu_i^k} - 1 \right)^2 \right\}. \quad (3.25)$$

A relação (3.25) impede de obter-se a equivalência com Afim Escala, pois o termo μ_i^k multiplica o termo quadrático $\left(\frac{\mu_i}{\mu_i^k} - 1 \right)^2$ em cada termo do somatório.

No próximo capítulo vamos abordar melhor a penalidade θ quadrática.

Capítulo 4

Penalidade Quadrática

É neste capítulo que apresentamos a principal contribuição deste trabalho. Para formar a penalidade p **tipo 1** ou **2**, será utilizada uma função real θ quadrática que não verifica todas as propriedades (i) à (v) dadas na definição (2.3.1), e portanto, as penalidades **tipo 1** ou **2** não serão da família \mathcal{P} .

Nesse caso, em que as penalidades são quadráticas, os multiplicadores gerados pelo algoritmo de Lagrangeano Aumentado podem ser negativos, pois a derivada da função não é crescente em toda a reta, como no caso da família \mathcal{P} .

Levando em consideração que alguns métodos utilizam um modelo quadrático da função original, como é o caso de Região de Confiança, mostraremos que a aproximação quadrática com as penalidades p **tipo 1** e **2**, para obtenção do ponto x^{k+1} do algoritmo de Lagrangeano Aumentado coincide com conjugada da quase-distância das penalidades p **tipo 1** e **2**.

No capítulo anterior mostramos uma certa relação de equivalência entre os métodos de Lagrangeano Aumentado e Afim Escala. Aproveitando o teorema (3.6.1) vamos mostrar um resultado referente à geometria das curvas de nível das conjugadas das penalidades **tipo 1** e **2** neste caso quadrático.

No final deste capítulo, apresentaremos uma heurística em que é possível contornar o problema de multiplicadores negativos encontrando um valor para o parâmetro r que torne $\mu^{k+1} \geq 0$, mas que foge às metodologias tradicionais de algoritmo de Lagrangeano Aumentado.

4.1 A Penalidade Quadrática Clássica

A função de penalidade proposta por Rockafellar [23] para o parâmetro de penalidade $r > 0$, tem a seguinte forma:

$$y \in \mathcal{R}, \mu \in \mathcal{R}_+ \mapsto p(y, \mu) = \frac{1}{2r} (\max \{0, ry + \mu\}^2 - \mu^2), \quad (4.1)$$

ou equivalentemente

$$p(y, \mu) = \begin{cases} \frac{r}{2}y^2 + \mu y, & \text{se } y \geq -\frac{\mu}{r} \\ -\frac{\mu^2}{2r}, & \text{se } y < -\frac{\mu}{r} \end{cases}. \quad (4.2)$$

Para $\mu \in \mathcal{R}_+$ fixo a derivada em (4.2) tem a forma

$$p'(y, \mu) = \begin{cases} ry + \mu, & \text{se } y \geq -\frac{\mu}{r} \\ 0, & \text{se } y < -\frac{\mu}{r} \end{cases}. \quad (4.3)$$

As figuras abaixo mostram o gráfico da função p (4.2) e da função p' (4.3) para o caso em que $r = \mu = 1$.

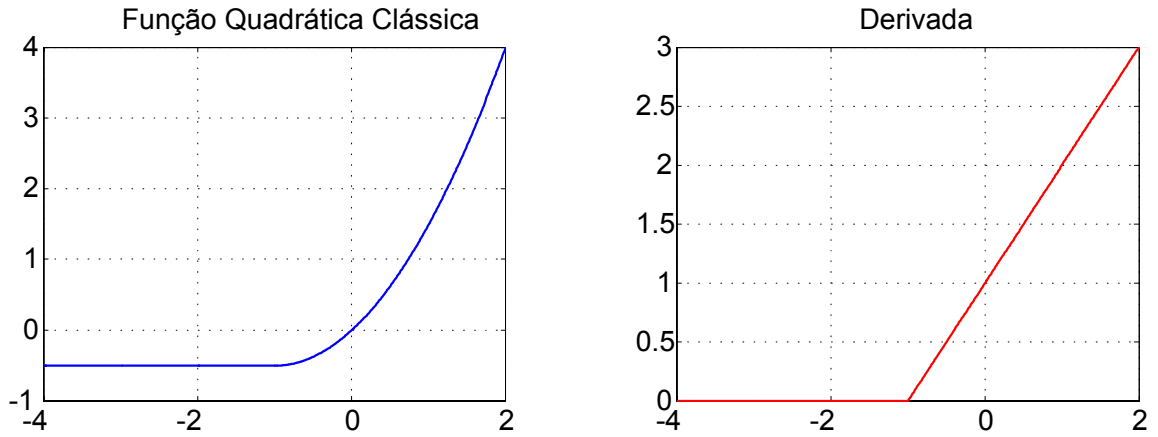


Figura 4.1: Função Quadrática Clássica

Como a função de Rockafellar não possui segunda derivada no ponto $y = -\frac{\mu}{r}$, podemos ter dificuldades ao usar métodos que utilizem derivada de ordem 2, como é o caso por exemplo, do método de Newton e o método de Região de Confiança.

Neste trabalho estamos propondo uma penalidade quadrática que possui derivada de segunda ordem, que apresentaremos nas próximas seções.

4.2 A Função Quadrática

A função θ a ser usada no algoritmo de Lagrangeano Aumentado é definida como

$$y \in \mathcal{R} \mapsto \theta(y) = \frac{1}{2}y^2 + y, \quad (4.4)$$

nota-se que (4.4) é uma aproximação de Taylor de ordem 2 das funções $\theta(y) = e^y - 1$ e $\theta(y) = -\ln(1 - y)$ no ponto $y = 0$. Sendo $\theta(y) = e^y - 1$ é utilizado por [1] e $\theta(y) = -\ln(1 - y)$ utilizado por [17].

A função conjugada dada na definição (1.9.1), ou seja,

$$s \in \mathcal{R}^n \mapsto f^*(s) = \sup_x \{x^T s - f(x), x \in \mathcal{R}^n\}, \quad (4.5)$$

para a conjugada da função θ dada na relação (4.4) é

$$s \in \mathcal{R} \mapsto \theta^*(s) = \frac{1}{2}(s - 1)^2. \quad (4.6)$$

As figuras abaixo mostram o gráfico da função θ (4.4) e da função θ^* (4.6).

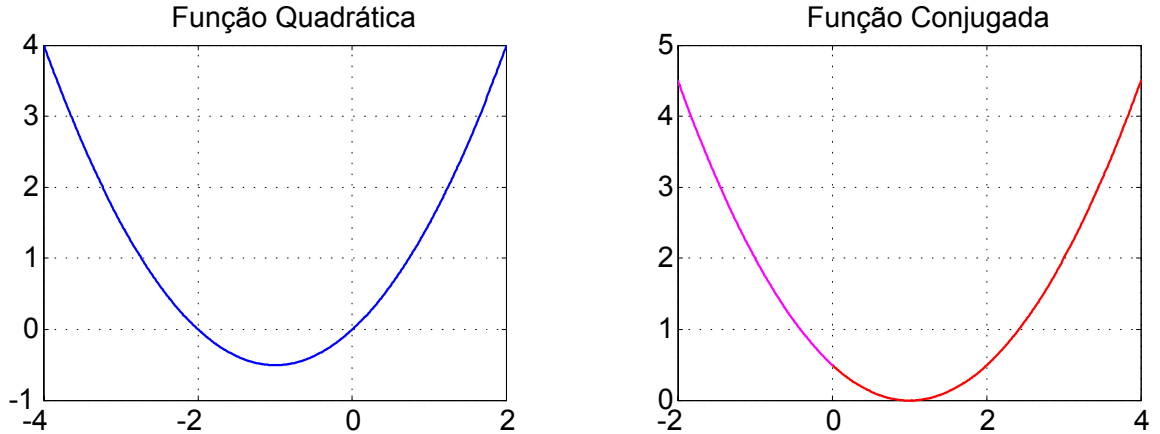


Figura 4.2: Quadrática e sua Conjugada

Observa-se que (4.4) não é crescente em todo seu domínio, isto é, no intervalo de $(-\infty, -1)$ é decrescente e no intervalo de $(-1, \infty)$ é crescente. Por não ser crescente em todo o seu domínio gera derivadas direcionais negativas. Com θ dada na relação (4.4) as

funções de penalidades construídas a partir dessa não serão da família \mathcal{P} . Observa-se que a propriedade de limitação por baixo da derivada, dada no item (P5) da definição (2.1.1), não é satisfeita aqui, pois $\lim_{y \rightarrow -\infty} \theta'(y) = -\infty$, que deveria ser $\lim_{y \rightarrow -\infty} \theta'(y) = 0$.

4.3 Funções de Penalidade p Quadráticas

Nesta seção faremos o estudo da penalidade p **tipo 1** e **2** dadas nas relações (2.3) e (2.5), respectivamente, com θ quadrática dada na relação (4.4)

Seja $\theta : \mathcal{R} \rightarrow \mathcal{R}$ definida por

$$\theta(y) = \frac{1}{2}y^2 + y. \quad (4.7)$$

A penalidade p **tipo 1** dada em relação (2.3), isto é,

$$y \in \mathcal{R}, \mu \in \mathcal{R}_{++} \mapsto p(y, \mu) = \theta(\mu y) \in \mathcal{R}, \quad (4.8)$$

juntamente com a função θ da relação (4.7) determina a seguinte função p quadrática

$$y \in \mathcal{R}, \mu \in \mathcal{R}_{++} \mapsto p(y, \mu) = \frac{1}{2}(\mu y)^2 + (\mu y). \quad (4.9)$$

Usando a relação (4.6), a conjugada de p da relação (4.9) é da forma

$$(s, \mu) \in \mathcal{R}_+ \times \mathcal{R}_{++} \mapsto p^*(s, \mu) = \frac{1}{2} \left(\frac{s}{\mu} - 1 \right)^2. \quad (4.10)$$

A conjugada de p **tipo 1** dada na relação (4.10) juntamente com o item (viii) da proposição (1.9.6) determinam uma quase-distância que tem a seguinte forma

$$(s, \mu) \in \mathcal{R}_+^m \times \mathcal{R}_{++}^m \mapsto D(s, \mu) = \sum_{i=1}^m \theta^* \left(\frac{s_i}{\mu_i} \right). \quad (4.11)$$

Agora se θ é a função quadrática (4.4), a relação (4.11) fica

$$(s, \mu) \in \mathcal{R}_+^m \times \mathcal{R}_{++}^m \mapsto D(s, \mu) = \sum_{i=1}^m \frac{1}{2} \left(\frac{s_i}{\mu_i} - 1 \right)^2. \quad (4.12)$$

A penalidade p **tipo 2** dada em relação (2.5), isto é,

$$(y, \mu) \in \mathcal{R} \times \mathcal{R}_{++} \mapsto p(y, \mu) = \mu\theta(y) \in \mathcal{R} \quad (4.13)$$

juntamente com a função θ da relação (4.7) determina a seguinte função p quadrática

$$y \in \mathcal{R}^m, \mu \in \mathcal{R}_{++}^m \mapsto p(y, \mu) = \frac{1}{2}\mu(y)^2 + \mu(y). \quad (4.14)$$

A conjugada de p , dada na relação (4.14) é da forma

$$(s, \mu) \in \mathcal{R}_+^m \times \mathcal{R}_{++}^m \mapsto p^*(s, \mu) = \frac{\mu}{2} \left(\frac{s}{\mu} - 1 \right)^2. \quad (4.15)$$

A conjugada de p **tipo 2** dada na relação (4.15), juntamente com o item (viii) da proposição (1.9.6), determinam uma quase-distância que tem a seguinte forma

$$(s, \mu) \in \mathcal{R}_+^m \times \mathcal{R}_{++}^m \mapsto D(s, \mu) = \sum_{i=1}^m \mu_i \theta^* \left(\frac{s_i}{\mu_i} \right). \quad (4.16)$$

Agora se θ é a função quadrática (4.4), a relação (4.16) fica

$$(s, \mu) \in \mathcal{R}_+^m \times \mathcal{R}_{++}^m \mapsto D(s, \mu) = \sum_{i=1}^m \frac{\mu_i}{2} \left(\frac{s_i}{\mu_i} - 1 \right)^2. \quad (4.17)$$

Levando em consideração que alguns métodos utilizam um modelo quadrático da função original, como é o caso de Região de Confiança, olharemos o que acontece com a aproximação quadrática das quase-distâncias (4.11) e (4.16), ou seja,

$$(s, \mu) \in \mathcal{R}_+^m \times \mathcal{R}_{++}^m \mapsto D(s, \mu) = \sum_{i=1}^m \theta^* \left(\frac{s_i}{\mu_i} \right)$$

e

$$(s, \mu) \in \mathcal{R}_+^m \times \mathcal{R}_{++}^m \mapsto D(s, \mu) = \sum_{i=1}^m \mu_i \theta^* \left(\frac{s_i}{\mu_i} \right).$$

Dado $\mu \in \mathcal{R}_{++}^m$, considere a aproximação quadrática de uma quase-distância D ,

próxima ao ponto \bar{s} , ou seja,

$$(s, \mu) \in \mathcal{R}_+^m \times \mathcal{R}_{++}^m \mapsto D(s, \mu) \sim D(\bar{s}, \mu) + \nabla D_s(\bar{s}, \mu)^T (s - \bar{s}) + \frac{1}{2} (s - \bar{s})^T \nabla^2 D_s(\bar{s}, \mu) (s - \bar{s}). \quad (4.18)$$

Para o caso (4.11), a aproximação quadrática (4.18) fica

$$D(s, \mu) \sim \sum_{i=1}^m \left[\theta^* \left(\frac{\bar{s}_i}{\mu_i} \right) + \frac{1}{\mu_i} \theta'^* \left(\frac{\bar{s}_i}{\mu_i} \right) (s_i - \bar{s}_i) + \frac{1}{2(\mu_i)^2} \theta''^* \left(\frac{\bar{s}_i}{\mu_i} \right) (s_i - \bar{s}_i)^2 \right]. \quad (4.19)$$

Substituindo a função θ^* da relação (4.6), em (4.19) tem-se

$$D(s, \mu) \sim \sum_{i=1}^m \left[\left(\frac{\bar{s}_i}{\mu_i} - 1 \right)^2 + \frac{1}{\mu_i} \left(\frac{\bar{s}_i}{\mu_i} - 1 \right) (s_i - \bar{s}_i) + \frac{1}{2(\mu_i)^2} (s_i - \bar{s}_i)^2 \right],$$

que agrupando convenientemente os termos e fazendo $\bar{s} = \mu$, tem-se

$$\sum_{i=1}^m \frac{1}{2} \left(\frac{s_i}{\bar{s}_i} - 1 \right)^2. \quad (4.20)$$

Repetindo-se os passos anteriores para a função dada na relação (4.16) tem-se

$$D(s, \mu) \sim \sum_{i=1}^m \left[\mu_i \left(\frac{\bar{s}_i}{\mu_i} - 1 \right)^2 + \left(\frac{\bar{s}_i}{\mu_i} - 1 \right) (s_i - \bar{s}_i) + \frac{1}{2(\mu_i)} (s_i - \bar{s}_i)^2 \right],$$

que agrupando convenientemente os termos e fazendo $\bar{s} = \mu$, tem-se

$$\sum_{i=1}^m \frac{\bar{s}_i}{2} \left(\frac{s_i}{\bar{s}_i} - 1 \right)^2. \quad (4.21)$$

Essa expressão difere de (4.20) pelo fator multiplicativo \bar{s}_i . A figura (4.3) mostra elipsóides gerados, respectivamente, por (4.20) e (4.21). O elipsóide externo é referente à função (4.20) e o interno referente à (4.21). Veja que o elipsóide gerado por (4.20) tem formato de elipsóides de Dikin.

Em seguida olharemos a geometria de curvas das quase-distâncias dadas pelas conjugadas das penalidades **tipo 1** e **2**, respectivamente, no caso em que θ é quadrática.

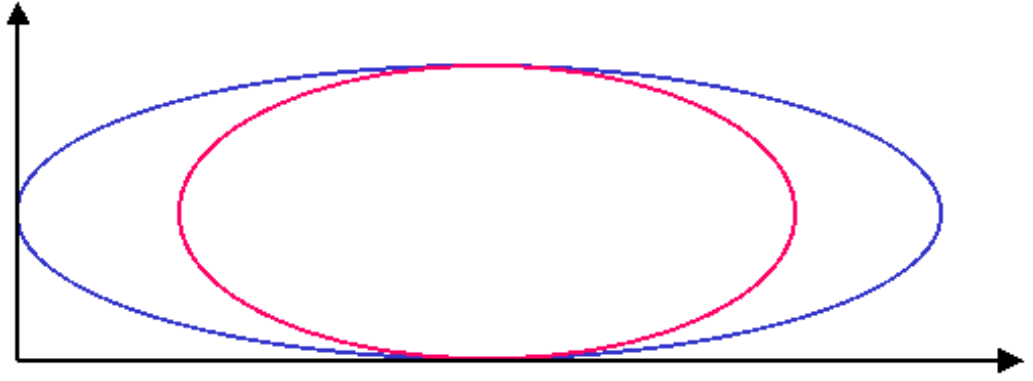


Figura 4.3: Curvas de Nível das Aproximações Quadráticas de Quase-distâncias

4.4 Geometria das Quase-distâncias dadas pelas Conjugadas das Penalidades Quadráticas.

As quase-distâncias dadas pelas conjugadas das penalidades p , **tipo 1** e **2** e a função θ^* dada na relação (4.6), são, respectivamente,

$$(s, \mu) \in \mathcal{R}_+^m \times \mathcal{R}_{++}^m \mapsto D(s, \mu) = \sum_{i=1}^m \frac{1}{2} \left(\frac{s_i}{\mu_i} - 1 \right)^2$$

e

$$(s, \mu) \in \mathcal{R}_+^m \times \mathcal{R}_{++}^m \mapsto D(s, \mu) = \sum_{i=1}^m \frac{\mu_i}{2} \left(\frac{s_i}{\mu_i} - 1 \right)^2.$$

Exemplo 4.4.1 A função $D: \mathcal{R}_+^2 \times \mathcal{R}_{++}^2 \rightarrow \mathcal{R}$, dada pela conjugada da penalidade **tipo 1** com θ quadrática é da forma

$$\mu \in \mathcal{R}_+^2, \mu^k \in \mathcal{R}_{++}^2 \mapsto D(\mu, \mu^k) = \left(\frac{\mu_1}{\mu_1^k} - 1 \right)^2 + \left(\frac{\mu_2}{\mu_2^k} - 1 \right)^2. \quad (4.22)$$

Exemplo 4.4.2 A função $D: \mathcal{R}_+^2 \times \mathcal{R}_{++}^2 \rightarrow \mathcal{R}$, dada pela conjugada da penalidade **tipo 2** com θ quadrática é da forma

$$\mu \in \mathcal{R}_+^2, \mu^k \in \mathcal{R}_{++}^2 \mapsto D(\mu, \mu^k) = \mu_1^k \left(\frac{\mu_1}{\mu_1^k} - 1 \right)^2 + \mu_2^k \left(\frac{\mu_2}{\mu_2^k} - 1 \right)^2. \quad (4.23)$$

A geometria das curvas de nível da função D para os exemplos acima (4.3.1) e (4.3.2), pode ser visto na figura (4.4). O elipsóide externo é referente à função dada na relação (4.22) e o interno à função dada na relação (4.23).

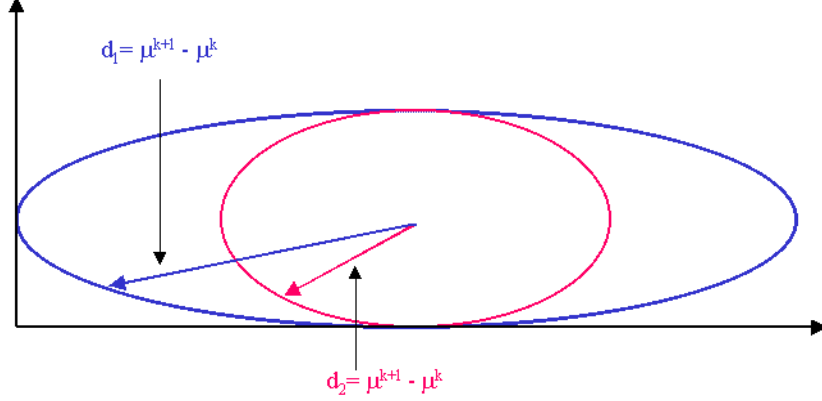


Figura 4.4: Direções Duais Geradas pelo Algoritmo de Lagrangeano Aumentado

Sejam as direções duais d_1 e d_2 geradas pelo algoritmo de Lagrangeano Aumentado quando aplicado respectivamente com as penalidades **tipo 1** e **2**. Veja que estamos usando a equivalência entre Ponto Proximal e Lagrangeano Aumentado. Assim, as curvas da figura (4.4) representam direções duais, pois usamos as curvas de nível das funções conjugadas.

Uma das vantagens de gerar direções equivalentes às direções geradas pelo algoritmo Afim Escala é que o comprimento do passo dado pelo algoritmo é bem mais longo. Veja que o formato do elipsóide formado pela função dada na relação (4.22) é bem mais alongado que o elipsóide da função dada na relação (4.23). No próximo teorema provado em [17] destacaremos esse fato, cuja demonstração reescreveremos para readaptando-a ao nosso trabalho.

Teorema 4.4.3 *Considere $\mu \in \mathcal{R}_+^m$ e $D: \mathcal{R}_+^m \times \mathcal{R}_{++}^m \rightarrow \mathcal{R}$ uma quase-distância quadrática, tais que seus conjuntos de nível determinados pela constante c , são representados por*

$$C = \{\mu \in \mathcal{R}_+^m : D(\mu, \mu^k) = c\}. \quad (4.24)$$

Se a função D é dada por

$$\mu \in \mathcal{R}_+^m, \mu^k \in \mathcal{R}_{++}^m \mapsto D(\mu, \mu^k) = \sum_{i=1}^m \frac{1}{2} \left(\frac{\mu_i}{\mu_i^k} - 1 \right)^2 \quad (4.25)$$

então o maior valor que a constante c pode assumir tal que $C \subset \mathcal{R}_+^m$ é $c = \frac{1}{2}$.

Se a função D é dada por

$$\mu \in \mathcal{R}_+^m, \mu^k \in \mathcal{R}_{++}^m \mapsto D(\mu, \mu^k) = \sum_{i=1}^m \frac{\mu_i^k}{2} \left(\frac{\mu_i}{\mu_i^k} - 1 \right)^2 \quad (4.26)$$

então o maior valor que a constante c pode assumir tal que $C \subset \mathcal{R}_+^m$ é $c = \frac{1}{2} \min\{\mu^k\}$.

Prova. Dados $\mu \in \mathcal{R}_+^m$ e $D : \mathcal{R}_+^m \times \mathcal{R}_{++}^m \rightarrow \mathcal{R}$ uma quase-distância quadrática, queremos determinar o maior valor de c tal que (4.24) esteja inteiramente contido no ortante positivo, ou seja, esteja inteiramente contido em \mathcal{R}_+^m .

Se D é dada, respectivamente, por (4.25) e (4.26), o conjunto C (4.24) representa elipsóides com eixos dados pelas coordenadas de μ^k que são paralelos aos eixos coordenados de \mathcal{R}^m .

O menor valor das coordenadas de $\mu^k = \min\{\mu^k\}$ determina o valor da constante c onde o elipsóide tangencia um dos eixos coordenados do \mathcal{R}^m . Para exemplificar, considere $\mu \in \mathcal{R}_+^2$ e a figura (4.5) a seguir.

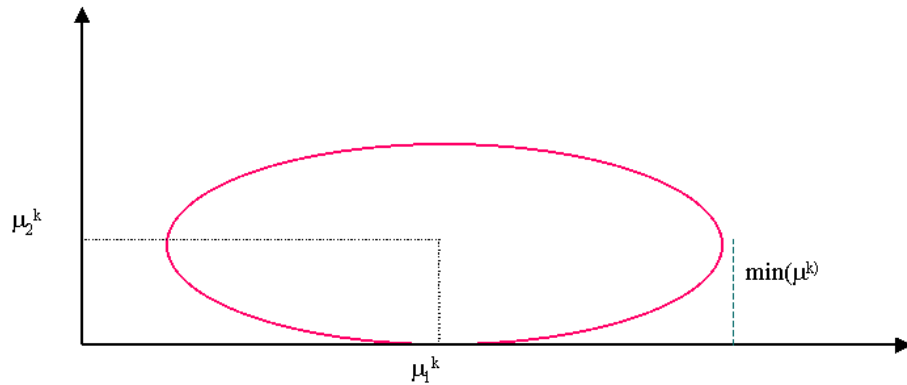


Figura 4.5: Coordenada Mínima

Para determinar o ponto de tangência, isto é, o valor da constante c , con-

sidere para algum $\ell \in \{1, 2, \dots, m\}$, $\mu_\ell^k = \min\{\mu^k\}$ e escolha $\mu = (\mu_1^k, \dots, \mu_{\ell-1}^k, 0, \mu_{\ell+1}^k, \dots, \mu_m^k)$. Se a função D é dada por (4.25), tem-se

$$c = \sum_{i=1, i \neq \ell}^m \frac{1}{2} \underbrace{\left(\frac{\mu_i}{\mu_i^k} - 1 \right)^2}_{=0} + \frac{1}{2} \underbrace{\left(\frac{0}{\mu_\ell^k} - 1 \right)^2}_{=1} = \frac{1}{2} \quad (4.27)$$

sendo $\mu = \mu_i^k$, tem-se $\left(\frac{\mu_i}{\mu_i^k} - 1 \right)^2 = 0$, para todo $i=1, 2, \dots, \ell-1, \ell+1, \dots, m$. Portanto $c=\frac{1}{2}$.

Da mesma forma se a função D é dada por (4.26), tem-se

$$c = \sum_{i=1, i \neq \ell}^m \frac{\mu_i^k}{2} \underbrace{\left(\frac{\mu_i}{\mu_i^k} - 1 \right)^2}_{=0} + \frac{\mu_\ell^k}{2} \underbrace{\left(\frac{0}{\mu_\ell^k} - 1 \right)^2}_{=1} = \frac{\mu_\ell^k}{2}. \quad (4.28)$$

Como $\mu_\ell^k = \min\{\mu^k\}$, tem-se $c=\frac{1}{2} \min\{\mu^k\}$. ■

Pelo teorema anterior, o maior elipsóide inteiramente contido em \mathcal{R}_+^m , para as funções dadas em (4.25) e (4.26) são, respectivamente, da forma

$$c = \sum_{i=1}^m \frac{1}{2} \left(\frac{\mu_i}{\mu_i^k} - 1 \right)^2 = \frac{1}{2}$$

e

$$c = \sum_{i=1}^m \frac{\mu_i^k}{2} \left(\frac{\mu_i}{\mu_i^k} - 1 \right)^2 = \frac{1}{2} \min\{\mu^k\}.$$

Da prova do teorema anterior, quando escolhemos μ com coordenada ℓ nula e as demais iguais às de μ^k , encontramos o maior elipsóide contido em \mathcal{R}_+^m . Esta informação será utilizada na próxima seção em que apresentaremos uma heurística para a atualização do parâmetro de penalidade r .

4.5 Uma Heurística sobre a Atualização do Parâmetro de Penalidade

Nesta seção descreveremos uma heurística, para a penalidade quadrática, sobre a atualização do parâmetro de penalidade nos algoritmos de Lagrangeano Aumentado.

Essa heurística surgiu dos resultados provados nos capítulos anteriores. No teorema (2.6.3) foi mostrada a equivalência entre Lagrangeano Aumentado e Ponto Proximal, já na proposição (2.7.1) foi exposta uma certa equivalência entre Ponto Proximal e Região de Confiança. Além disso, foi mostrado no lema (2.7.2) que aumentando-se o parâmetro de penalidade no algoritmo de Ponto Proximal diminui-se a região no método de Região de Confiança. Dessa forma, no caso de programação linear, existe um $\Delta > 0$ (raio da Região de Confiança) que determina o maior elipsóide no algoritmo de Região de Confiança (ver capítulo 5, seção 5.2) que fica totalmente no ortante positivo. Pela equivalência entre Região de Confiança e Ponto Proximal, dada na proposição (2.7.1), existe um r (parâmetro de penalidade) no algoritmo de Ponto Proximal, tal que, a solução desses dois métodos coincide e fica no ortante positivo.

A seguir, reescreveremos o algoritmo de Lagrangeano Aumentado, aplicado ao problema linear, com penalidade quadrática. Neste caso, em que a penalidade não é da família \mathcal{P} , os multiplicadores gerados podem ser negativos. Após o algoritmo veremos como contornar esse problema.

Algoritmo 4.5.1 Dados $\mu^0 > 0$, $r^0 > 0$.

faça $k = 0$

repita

encontre

$$x^+ \in \arg \min_{x \in \mathbb{R}^n} \left\{ -b^T x + \sum_{i=1}^m r^k p \left(\frac{A_i(x) - c_i}{r^k}, \mu_i^k \right) \right\}$$

calcule

$$\mu_i^+ = p' \left(\frac{A_i(x^+) - c_i}{r^k}, \mu_i^k \right), \quad i = 1, 2, \dots, m$$

se $\mu^+ \geq 0$

$$x^{k+1} = x^+$$

$$\mu^{k+1} = \mu^+$$

$$r^{k+1} < r^k$$

senão

$$r^{k+1} = -\min\{\mu_i \cdot g_i(x^+), \quad (\text{para } p \text{ tipo 1})$$

$$\mu_i^{k+1} = p' \left(\frac{A_i(x^+) - c_i}{r^{k+1}}, \mu_i^k \right), \quad i = 1, 2, \dots, m$$

$$k = k + 1$$

Cada iteração do algoritmo (4.5.1) consiste em formar subproblemas irrestritos, que podem ser resolvidos por qualquer método de minimização irrestrita, e atualizar parâmetros. Se $\mu^+ \geq 0$, atualiza-se o ponto corrente e o multiplicador de Lagrange para a próxima iteração, reduz-se o parâmetro r de penalidade. Caso contrário, isto é, $\mu^+ \not\geq 0$, mantém-se o ponto e calcula-se o valor do parâmetro de penalidade e atualiza-se o vetor de multiplicadores de Lagrange. Esta parte é nova e foge à filosofia dos métodos de Lagrangeano Aumentado. Por isso estamos tratando, esta seção, como heurística. Os resultados numéricos alcançados com essa heurística foram superiores à metodologia tradicional.

A seguir, vamos apresentar resultados que serão utilizados posteriormente nas implementações do algoritmo de Lagrangeano Aumentado.

Considere na proposição (4.5.2) a seguir, $g_i(x^+) = A_i(x^+) - c_i$ para $i = 1, 2, \dots, m$.

Proposição 4.5.2 *Considere x^+ e μ^+ gerados pelo algoritmo de Lagrangeano Aumentado (4.5.1) aplicado ao problema linear com penalidade **tipo 1** quadrática (4.4), se $\mu^+ \not\geq 0$ e o parâmetro de penalidade $r = -\min_i \{\mu_i^+ \cdot g_i(x^+)\}$, então, tem-se $\mu_i^{k+1} = p' \left(\frac{g_i(x^+)}{r^{k+1}}, \mu_i^k \right) \geq 0$.*

Prova. *Sejam x^+ e μ^+ gerados pelo algoritmo de Lagrangeano Aumentado (4.5.1). Se $\mu^+ \geq 0$, faz-se $\mu^{k+1} = \mu^+$, mas quando μ^+ (multiplicador de Lagrange) tiver coordenadas negativas, contraria as condições de K.K.T. dada na relação (2.10).*

No algoritmo (4.5.1) do método de Lagrangeano Aumentado podemos notar que o valor de μ^+ é atualizado pela fórmula

$$\mu_i^+ = p' \left(\frac{g_i(x^+)}{r^k}, \mu_i^k \right) \quad \text{para } i = 1, 2, \dots, m. \quad (4.29)$$

*Como vamos utilizar a penalidade **tipo 1** com θ quadrática (4.4), isto é, $p(y, \mu) = \theta(\mu y)$, tem-se a seguinte forma para p*

$$y \in \mathcal{R}, \mu \in \mathcal{R}_{++} \rightarrow p(y, \mu) = \frac{1}{2}(\mu y)^2 + (\mu y)$$

e p' é da seguinte forma

$$y \in \mathcal{R}, \mu \in \mathcal{R}_{++} \rightarrow p'(y, \mu) = \mu \theta'(y\mu) = \mu(\mu y + 1).$$

Dessa forma, a atualização do multiplicador de Lagrange (4.29) é dada por

$$\mu_i^+ = \mu_i^k \left(\frac{\mu_i^k \cdot g_i(x^+)}{r^k} + 1 \right) \quad \text{para } i = 1, 2, \dots, m.$$

e sabendo que μ_i^k é positivo para $i = 1, 2, \dots, m$, se $u^+ \not\geq 0$ tem-se

$$\mu_i^+ = \underbrace{\mu_i^k}_{\geq 0} \left(\underbrace{\frac{\mu_i^k \cdot g_i(x^+)}{r^k} + 1}_{\leq 0} \right). \quad (4.30)$$

De (4.30) tem-se $g_i(x^+) < 0$, pois $r^k > 0$. Seja $\ell \in \{1, 2, \dots, m\}$ e considere $\mu_\ell^+ \cdot g_\ell(x^+) = \min\{\mu_i^k \cdot g_i(x^+)\}$. Escolhendo o valor do parâmetro de penalidade

$$r^{k+1} = -\mu_\ell^k \cdot g_\ell(x^+) \quad (4.31)$$

tem-se

$$\mu_\ell^{k+1} = \underbrace{\mu_\ell^k}_{\geq 0} \left(\underbrace{\frac{\mu_\ell^k \cdot g_\ell(x^+)}{-\mu_\ell^k \cdot g_\ell(x^+)} + 1}_{= 0} \right) = 0. \quad (4.32)$$

Em seguida mostramos que as demais componentes de μ^{k+1} são não negativas.

Como $r^{k+1} = -\mu_\ell^k \cdot g_\ell(x^+)$ e $g_\ell(x^+) < 0$ tem-se:

i) Se $g_\ell(x^+) < 0$ e $\mu_\ell^k = 0$ em (4.32) $\mu_\ell^{k+1} = \mu_\ell^k \geq 0$.

ii) Se $g_\ell(x^+) < 0$ e $\mu_\ell^k > 0$, então $r^{k+1} > 0$, além disso, r^{k+1} foi tomado como a maior componente do vetor $[\mu_1^k \cdot g_1(x^+), \mu_2^k \cdot g_2(x^+), \dots, \mu_m^k \cdot g_m(x^+)]^T$ então $\mu_i^{k+1} = \mu_i^k \left(\frac{\mu_i^k \cdot g_i(x^+)}{r^{k+1}} + 1 \right) \geq 0$, para $i = 1, 2, \dots, m$. ■

A seguir, apresentaremos um exemplo para fixar a idéia da proposição acima. No exemplo (4.5.4) usaremos a notação (4.5.3) dada a seguir.

Notação 4.5.3 O produto $\mu^k \cdot g(x^+)$ é componente a componente.

Exemplo 4.5.4 Seja $\mu^k = \begin{pmatrix} 1 \\ 2 \\ 2 \\ 1 \end{pmatrix}$ e $g(x^+) = \begin{pmatrix} 2 \\ -1 \\ -2 \\ 1 \end{pmatrix}$ e $r^k = 1$, a atualização de μ^+ , utilizando a penalidade **p tipo 1** com a função θ quadrática (4.4) dada na relação (4.29), fica

$$\mu^+ = \mu^k \left(\frac{\mu^k \cdot g(x^+)}{r^k} + e \right), \quad e = (1, 1, \dots, 1)^T \quad (4.33)$$

ou

$$\mu^+ = \begin{pmatrix} 3 \\ -2 \\ -6 \\ 2 \end{pmatrix}$$

como $\mu^+ \not\geq 0$ mantém-se, para a próxima iteração o ponto x^+ , e

$$\mu^k \cdot g(x^+) = \begin{pmatrix} 2 \\ -2 \\ -4 \\ 1 \end{pmatrix}.$$

Escolhe o valor do parâmetro r^{k+1} dado na relação (4.31), isto é, $r = -\min(\mu^k \cdot g(x^+)) = -(-4)=4$, logo em (4.33), fica

$$\mu^+ = \begin{pmatrix} 1 \\ 2 \\ 2 \\ 1 \end{pmatrix} \begin{pmatrix} \frac{1*2}{4} + 1 \\ \frac{2*(-1)}{4} + 1 \\ \frac{2*(-2)}{4} + 1 \\ \frac{1*1}{4} + 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 2 \\ 1 \end{pmatrix} \begin{pmatrix} 1.5 \\ 0.5 \\ 0 \\ 1.25 \end{pmatrix} = \begin{pmatrix} 1.5 \\ 1 \\ 0 \\ 1.25 \end{pmatrix} \geq 0.$$

Considere na proposição (4.5.5) a seguir, $g_i(x^+) = A_i(x^+) - c_i$ para $i = 1, 2, \dots, m$.

Proposição 4.5.5 Considere x^+ e μ^+ gerados pelo algoritmo de Lagrangeano Aumentado (4.5.1) aplicado ao problema linear com penalidade p **tipo 2** quadrática (4.4), se $\mu^+ \not\geq 0$ e o parâmetro de penalidade $r = -\min_i \{u_i^+ \cdot g_i(x^{k+1})\}$, então, tem-se $\mu^{k+1} = p' \left(\frac{g_i(x^+)}{r^{k+1}}, \mu_i^k \right) \geq 0$.

Prova. Sejam x^+ e μ^+ gerados pelo algoritmo de Lagrangeano Aumentado (4.5.1). Se $\mu^+ \geq 0$, faz-se $\mu^{k+1} = \mu^+$, mas quando μ^+ (multiplicador de Lagrange) tiver coordenadas negativas, contraria as condições de K.K.T. dada na relação (2.10).

No algoritmo (4.5.1) do método de Lagrangeano Aumentado podemos notar que o valor de μ^+ é atualizado pela fórmula

$$\mu_i^+ = p' \left(\frac{g_i(x^+)}{r^k}, \mu_i^k \right) \quad \text{para } i = 1, 2, \dots, m. \quad (4.34)$$

Como vamos utilizar a penalidade **tipo 2** com θ quadrática (4.4), isto é, $p(y, \mu) = \mu\theta(y)$, tem-se a seguinte forma para p

$$y \in \mathcal{R}, \mu \in \mathcal{R}_{++} \rightarrow p(y, \mu) = \frac{\mu}{2}(y)^2 + (\mu y)$$

e p' é da seguinte forma

$$y \in \mathcal{R}, \mu \in \mathcal{R}_{++} \rightarrow p'(y, \mu) = \mu\theta'(y) = \mu(y + 1).$$

Dessa forma, a atualização do multiplicador de Lagrange (4.34) é dada por

$$\mu_i^+ = \mu_i^k \left(\frac{g_i(x^+)}{r^k} + 1 \right) \quad \text{para } i = 1, 2, \dots, m. \quad (4.35)$$

e sabendo que μ_i^k é positivo para $i = 1, 2, \dots, m$, se $u^+ \not\geq 0$ tem-se

$$\mu_i^+ = \underbrace{\mu_i^k}_{\geq 0} \left(\underbrace{\frac{g_i(x^+)}{r^k} + 1}_{\leq 0} \right). \quad (4.36)$$

De (4.36) tem-se $g_i(x^+) < 0$, pois $r^k > 0$. Seja $\ell \in \{1, 2, \dots, m\}$ e considere

$g_\ell(x^+) = \min\{g_i(x^+)\}$. Escolhendo o valor do parâmetro de penalidade

$$r^{k+1} = -g_\ell(x^+) \quad (4.37)$$

tem-se

$$u_\ell^{k+1} = \underbrace{\mu_\ell^k}_{\geq 0} \left(\underbrace{\frac{g_\ell(x^+)}{-g_\ell(x^+)} + 1}_{= 0} \right) = 0. \quad (4.38)$$

Em seguida mostramos que as demais componentes de μ^{k+1} são não negativas.

Como $r^{k+1} = -g_\ell(x^+)$ e $g_\ell(x^+) < 0$, tem-se que $g_\ell(x^+) < 0$ e $\mu_\ell^k > 0$, então $r^{k+1} > 0$, além disso, r^{k+1} foi tomado como a maior componente do vetor $[g_1(x^+), g_2(x^+), \dots, g_m(x^+)]^T$ então $\mu_i^{k+1} = \mu_i^k \left(\frac{\mu_i^k \cdot g_i(x^+)}{r^{k+1}} + 1 \right) \geq 0$, para $i = 1, 2, \dots, m$. ■

Na implementação do algoritmo de Lagrangeano Aumentado, descrita no próximo capítulo, serão utilizadas as informações das proposições (4.5.2) e (4.5.5) para a atualização do parâmetro de penalidade. Como os multiplicadores serão atualizados após o parâmetro de penalidade, estamos gerando, nesse caso, algoritmos de Lagrangeano Aumentado que fogem às metodologias usuais. Por essa razão, estamos considerando como uma heurística, pois nenhuma teoria de convergência foi feita para essa situação. No entanto, os resultados numéricos apresentados foram superiores àqueles apresentados com metodologias usuais, como será visto através dos resultados numéricos apresentados no Capítulo 6.

No próximo capítulo, apresentaremos a implementação do algoritmo de Lagrangeano Aumentado para as penalidades **p tipo 1** e **2**.

Capítulo 5

Implementação

Neste capítulo, tratamos da implementação dos algoritmos de Lagrangeano Aumentado. A função do parâmetro de penalidade é penalizar, então, se a penalização for muito rigorosa, os subproblemas gerados podem ser difíceis de resolver, mas por outro lado se a penalização for suave, podem ocorrer subproblemas fáceis mas serem necessárias muitas iterações na solução do problema. O ideal é ter uma boa fórmula para atualizar esse parâmetro, que, no entanto, não é fácil de conseguir.

5.1 Parâmetro de Penalidade

Aumento do Parâmetro r

Como a função de penalidade que estamos utilizando é a quadrática (4.4), a qual não pertence à família de penalidade \mathcal{P} , e a sua conjugada (4.6) não é coerciva na fronteira de \mathcal{R}_+^m , o algoritmo de Ponto Proximal pode gerar pontos não positivos, o mesmo pode ocorrer com o algoritmo de Lagrangeano Aumentado quando usa-se a função penalidade θ dada na relação (4.4).

No lema (2.7.2), mostramos que aumentando-se (diminuindo-se) o parâmetro r , diminui-se (aumenta-se) a Região de Confiança. Então, usando esse resultado vamos aumentar o valor do parâmetro r para diminuir a Região de Confiança forçando os multiplicadores a ficarem no ortante positivo.

Neste trabalho, o aumento do parâmetro r está relacionado com o problema de

multiplicadores negativos, isto é, $u_i^+ < 0$ para algum $i = 1, 2, \dots, m$, no algoritmo (3.4.1).

Para penalidade **tipo 1** e **2**, o valor do parâmetro r^{k+1} pode ser aumentado, multiplicando-o por um escalar $\gamma > 1$, isto é,

$$r^{k+1} = \gamma \cdot r^k. \quad (5.1)$$

Na heurística para contornar o problema de multiplicadores negativos, serão apresentadas duas maneiras de atualizar o valor de r :

- Se utilizarmos a penalidade **tipo 1** o parâmetro r pode ser atualizado utilizando o valor do r dado na proposição (4.4.2), isto é,

$$r = -\min_i \{u_i^+ \cdot g_i(x^+)\}; \quad (5.2)$$

- Se utilizarmos a penalidade **tipo 2** o parâmetro r pode ser atualizado utilizando o valor do r dado na proposição (4.4.5), isto é,

$$r = -\min_i \{g_i(x^+)\}. \quad (5.3)$$

Diminuição do Parâmetro r

Para motivar a necessidade de reduzir r^k em algoritmos com penalidade p **tipo 1**, apresentaremos no lema a seguir que se o parâmetro r é mantido constante a velocidade de convergência será sublinear.

Lema 5.1.1 *Considere (μ^k) uma seqüência gerada pelo algoritmo de Ponto Proximal (3.3.1) com quase-distância dada pela conjugada de p **tipo 1**, ou seja,*

$$(\mu, \mu^k) \in \mathcal{R}_+^m \times \mathcal{R}_{++}^m \rightarrow D(s, \mu) = \sum_{i=1}^m \theta^* \left(\frac{\mu_i}{\mu_i^k} \right),$$

e $r^k > \bar{r} > 0$, para todo k . Suponha que $\mu^k \rightarrow \bar{\mu}$ quando $k \rightarrow \infty$ e que exista $\ell \in \{1, 2, \dots, m\}$, tal que $\bar{\mu}_\ell = 0$. A seqüência (μ^k) gerada nessas condições tem convergência sublinear.

Prova. Ver [17] ■

O lema a seguir garante que a seqüência de multiplicadores gerada pelo algoritmo de Lagrangeano Aumentado (3.4.1) aplicado ao problema linear com penalidade p **tipo 2** tem convergência superlinear quando o parâmetro de penalidade tende a zero.

Lema 5.1.2 *Seja (μ^k) a seqüência gerada pelo algoritmo de Lagrangeano Aumentado (3.4.1) com parâmetro de penalidade $r=\frac{1}{c}$, então (μ^k) converge para a solução ótima dual. Além disso, se $c \rightarrow \infty$ então a convergência é superlinear.*

Prova. Ver [25] ■

A redução do parâmetro r é a mesma para as duas penalidades **tipo 1** e **2**. Se os multiplicadores são não negativos reduzimos o parâmetro r , da seguinte forma

$$r^{k+1} = \frac{r^k}{\alpha}$$

onde $\alpha > 1$ é um valor constante.

5.2 Algoritmo para o método de Região de Confiança

Aplicaremos o método de Região de Confiança em problemas irrestritos.

$$\begin{aligned} \min f(x) \\ x \in \mathcal{R} \end{aligned} \quad \text{(I)} \quad (5.4)$$

onde f é uma função de classe C^2 .

Em cada iteração k do algoritmo de Região de Confiança constrói-se um modelo quadrático da função objetivo, que é a aproximação de Taylor de ordem 2 em torno de x^k , e minimiza-se o modelo em uma certa região, ou seja

$$\begin{aligned} \min_{p \in \mathbb{R}^n} m_k(p) &= f_k + \nabla f_k^T p + \frac{1}{2} p^T B_k p \\ \text{sa : } \|p\| &\leq \Delta_k \end{aligned} \quad (5.5)$$

onde $f_k = f(x^k)$, $\nabla f_k = \nabla f(x^k)$ e $B_k \in \mathcal{R}^{n \times n}$ é simétrica.

A seguir apresentaremos o algoritmo de Região de Confiança [19].

Algoritmo 5.2.1 Dados $\overline{\Delta} > 0$, $\Delta_k \in (0, \overline{\Delta})$, e $\eta \in [0, \frac{1}{4}]$

faça $k = 0$

obtenha

$$p^k \in \arg \min \{f_k + \nabla f_k^T p + \frac{1}{2} p^T B_k p : \|p\| \leq \Delta_k\}$$

calcule

$$\rho_k = \frac{f(x^k) - f(x^k + p^k)}{m_k(0) - m_k(p^k)}$$

se $\rho_k < \frac{1}{4}$

$$\Delta_{k+1} = \frac{1}{4} \Delta_k$$

senão

se $\rho_k > \frac{3}{4}$ e $\|p\| = \Delta_k$

$$\Delta_{k+1} = \min(2\Delta_k, \overline{\Delta})$$

senão

$$\Delta_{k+1} = \Delta_k;$$

se $p > \eta$

$$x^{k+1} = x^k + p^k$$

senão

$$x^{k+1} = x^k$$

$k = k + 1$

Na iteração k do algoritmo resolve-se um subproblema, que consiste em minimizar uma função quadrática em uma certa região. Para a próxima iteração, se $f(x^k + p^k) < f(x^k)$. Toma-se $x^{k+1} = x^k + p^k$, onde p^k é a solução do subproblema quadrático na Região de Confiança, caso contrário $x^{k+1} = x^k$. A atualização de Δ_{k+1} depende de

$$\rho_k = \frac{f(x^k) - f(x^k + p^k)}{m_k(0) - m_k(p^k)}. \quad (5.6)$$

- Se $\rho_k < 0.25$ o modelo está ruim, nesse caso diminui-se o valor de Δ_{k+1} ;
- Se $\rho_k > 0.75$ o modelo está bom e Δ_{k+1} será aumentado sempre que p^k estiver na fronteira da região;
- Se $0.25 < \rho_k < 0.75$ o modelo está razoável e Δ_{k+1} é mantido como na prévia iteração.

O próximo teorema caracteriza quando o problema (5.5) tem solução.

Teorema 5.2.2 *O vetor p^* é uma solução global do problema de Região de Confiança*

$$\begin{aligned} \min_{p \in \mathbb{R}^n} m(p) &= f + \nabla f^T p + \frac{1}{2} p^T B p \\ \text{s.a.} \quad & \|p^*\| \leq \Delta \end{aligned}$$

se, e somente se p^* é viável e existe um escalar $\lambda \geq 0$, tal que:

$$\begin{aligned} (B + \lambda I)p^* &= -\nabla f \\ \lambda(\Delta - \|p^*\|) &= 0 \\ (B + \lambda I) &\geq 0 \end{aligned} \tag{5.7}$$

Prova. [19] ■

Baseado no teorema (5.2.2) tem-se o seguinte algoritmo (ver [19])

Algoritmo 5.2.3 *Dados $\lambda^0, \Delta > 0$:*

faça $k = 0$

fatore

$$(B + \lambda^k I) = R^T R$$

resolva

$$R^T R p^k = -\nabla f_k;$$

$$R^T q^k = p^k$$

atualize

$$\lambda^{(k+1)} = \lambda^{(k)} + \left(\frac{\|p^k\|}{\|q^k\|} \right)^2 \cdot \left(\frac{\|p^k\| - \Delta}{\Delta} \right)$$

$k = k + 1$

5.3 Algoritmos Implementados para o Método de Lagrangeano Aumentado

Nesta seção apresentaremos os quatro algoritmos de Lagrangeano Aumentado, que foram implementados neste trabalho. A implementação foi realizada utilizando o software Matlab versão 6.0.0.88 (R12).

Algoritmo de Lagrangeano Aumentado com penalidades p tipo 1 e 2 e θ quadrática

A função θ a ser utilizada para esta versão do algoritmo de Lagrangeano Aumentado é dada por (ver capítulo 4)

$$y \in \mathcal{R} \rightarrow \theta(y) = \frac{1}{2}y^2 + y. \quad (5.8)$$

Algoritmo 5.3.1 *Atualização do parâmetro de penalidade utilizando um escalar $\gamma > 1$ para penalidades p tipo 1 e 2 com θ quadrática*

Dados $\mu^0 > 0$, $\gamma > 1$, $\alpha > 1$ e $r^0 > 0$

faça $k = 0$

repita

encontre

$$x^+ \in \arg \min_{x \in \mathcal{R}^n} \left\{ f(x) + r^k \sum_{i=1}^m p \left(\frac{g_i(x)}{r^k}, \mu_i^k \right) \right\}$$

atualize

$$\mu_i^+ = p' \left(\frac{g_i(x^+)}{r^k}, \mu_i^k \right) \quad i = 1, 2, \dots, m$$

se $\mu_i^+ \geq 0$

$$x^{k+1} = x^+$$

$$\mu^{k+1} = \mu^+$$

$$r^{k+1} = \frac{r^k}{\alpha}$$

senão

$$r^{k+1} = \gamma \cdot r^k$$

$k = k + 1$

Cada iteração do algoritmo (5.3.1) consiste em formar subproblemas irrestritos, que podem ser resolvidos por qualquer método de minimização irrestrita. Para resolver o subproblema

$$x^+ \in \arg \min_{x \in \mathcal{R}^n} \left\{ f(x) + r^k \sum_{i=1}^m p \left(\frac{g_i(x)}{r^k}, \mu_i^k \right) \right\}$$

foi implementado o método de Região de Confiança baseado no algoritmo Hook em [4] e (ver também capítulo 4 de [19]), e atualiza-se os parâmetros:

- Se $\mu^+ \geq 0$, atualiza-se o ponto corrente e o multiplicador de Lagrange para a próxima iteração, reduz-se o parâmetro de penalidade r ;
- Se $\mu^+ \not\geq 0$, mantém-se o ponto, aumenta-se o valor do parâmetro de penalidade r .

Algoritmo de Lagrangeano Aumentado com penalidades p tipo 1 e 2 e θ a função m2b

A função θ a ser utilizada para esta versão do algoritmo de Lagrangeano Aumentado é dada por

$$y \in \mathcal{R} \rightarrow \theta(y) = \begin{cases} -\log(1-y) & \text{se } y \leq \frac{1}{2} \\ 2y^2 + \log(2) - \frac{1}{2} & \text{se } y > \frac{1}{2} \end{cases} \quad (5.9)$$

essa função é da família \mathcal{P} (ver [17]), por isso não gera multiplicadores negativos.

Algoritmo 5.3.2 Penalidades **tipo 1 e 2** e θ dada por (5.9)

Dados $\mu^0 > 0$, $\alpha > 1$ e $r^0 > 0$

faça $k = 0$

repita

encontre

$$x^+ \in \arg \min_{x \in \mathcal{R}^n} \left\{ f(x) + r^k \sum_{i=1}^m p \left(\frac{g_i(x)}{r^k}, \mu_i^k \right) \right\}$$

atualize

$$\mu_i^+ = p' \left(\frac{g_i(x^+)}{r^k}, \mu_i^k \right) \quad i = 1, 2, \dots, m$$

$$r^k = \frac{r^k}{\alpha}$$

$$k = k + 1$$

A metodologia dos algoritmos (5.3.1) e (5.3.2) é muito parecida. Porém o algoritmo (5.3.2) não gera multiplicadores negativos, isto é, $\mu^{k+1} \geq 0$.

Em seguida, apresentaremos dois algoritmos com penalidades que não pertencem a família \mathcal{P} , conforme a heurística apresentada no Capítulo 4.

Heurística para a atualização do parâmetro de penalidade no algoritmo de Lagrangeano Aumentado com θ quadrática

Enunciaremos a seguir o algoritmo implementado, com θ quadrática. A diferença em usarmos penalidade p **tipo 1** ou **2**, consiste na regra de atualização do parâmetro, como veremos no algoritmo a seguir.

Algoritmo 5.3.3 *Atualização do parâmetro de penalidade utilizando a heurística com penalidade p **tipo 1** ou **2** e θ quadrática*

Dados $\mu^0 > 0$, $\alpha > 1$, $\gamma > 1$ ($\alpha \neq \gamma$) e $r^0 > 0$

faça $k = 0$

repita

encontre

$$x_i^+ \in \arg \min_{x \in \mathcal{R}^n} \left\{ f(x) + r^k \sum_{i=1}^m p \left(\frac{g_i(x)}{r^k}, \mu_i^k \right) \right\}$$

atualize

$$\mu_i^+ = p' \left(\frac{g_i(x^+)}{r^k}, \mu_i^k \right) \quad i = 1, 2, \dots, m$$

se $\mu_i^+ \geq 0$

$$x^{k+1} = x^+$$

$$\mu^{k+1} = \mu^+$$

$$r^{k+1} = \frac{r^k}{\alpha}$$

senão

se $k = 0$

$$r^{k+1} = -\min\{\mu_i^k \cdot g_i(x^+)\}, \quad (\text{se } p \text{ é tipo 1})$$

$$r^{k+1} = -\min\{g_i(x^+)\}, \quad (\text{se } p \text{ é tipo 2})$$

$$\mu_i^{k+1} = p' \left(\frac{g_i(x^+)}{r^{k+1}}, \mu_i^k \right) \quad i = 1, 2, \dots, m$$

se $k \neq 0$

$$r^{k+1} = \gamma \cdot r^k$$

$k = k + 1$

Cada iteração do algoritmo (5.3.3) consiste em formar subproblemas irrestritos, que podem ser resolvidos por qualquer método de minimização irrestrita. Para resolver o

subproblema

$$x^+ \in \arg \min_{x \in \mathcal{R}^n} \left\{ f(x) + r^k \sum_{i=1}^m p \left(\frac{g_i(x)}{r^k}, \mu_i^k \right) \right\}$$

foi implementado o método de Região de Confiança baseado no algoritmo Hook em [4] e (ver também capítulo 4 de [19]), e atualiza-se os parâmetros:

- Se $\mu^+ \geq 0$, atualiza-se o ponto corrente e o multiplicador de Lagrange para a próxima iteração, reduz-se o parâmetro de penalidade r ;
- Se $\mu^+ \not\geq 0$, mantém-se o ponto, atualizam-se o valor do parâmetro de penalidade r e o novo multiplicador de Lagrange para a próxima iteração.

No próximo capítulo mostraremos os resultados obtidos com a penalidades p **tipo 1** e **2** e:

- θ quadrática dada na relação (5.8);
- θ m2b dada na relação (5.9).

Capítulo 6

Testes Numéricos

Neste capítulo vamos testar os algoritmos de Lagrangeano Aumentado cujas implementações foram descritas no Capítulo 5. Programamos em Matlab, versão 6.0.0.88 (R12) para Linux. O método de Região de Confiança é utilizado para resolver os subproblemas internos gerados em cada iteração do mesmo.

As penalidades utilizadas nos algoritmos implementados são:

$$(y, \mu) \in \mathcal{R} \times \mathcal{R}_{++} \rightarrow p(y, \mu) = \theta(\mu y) \in \mathcal{R} \cup \{+\infty\} \quad (6.1)$$

e

$$(y, \mu) \in \mathcal{R} \times \mathcal{R}_{++} \rightarrow p(y, \mu) = \mu\theta(y) \in \mathcal{R} \cup \{+\infty\} \quad (6.2)$$

e foram apresentadas anteriormente na seção (2.3), com as seguintes funções θ :

$$y \in \mathcal{R} \rightarrow \theta(y) = \frac{1}{2}y^2 + y, \quad (6.3)$$

e

$$y \in \mathcal{R} \rightarrow \theta(y) = \begin{cases} -\log(1-y) & \text{se } y \leq \frac{1}{2} \\ 2y^2 + \log(2) - \frac{1}{2} & \text{se } y > \frac{1}{2} \end{cases}. \quad (6.4)$$

Os problemas testados são da coleção CUTE: Constrained and Unconstrained Testing Environment [13], na versão small com interface para o Matlab. Por ser um

conjunto com milhares de problemas, escolhemos aqueles que satisfazem nossas condições, isto é, problemas com restrições de desigualdade e sem limites nas variáveis.

Testamos 82 (oitenta e dois) problemas que estão apresentados na tabela (6.1) dada a seguir. A primeira coluna da tabela é formada pelos nomes dos problemas apresentados na coleção CUTE. As demais colunas são: números de variáveis, restrições, tipo de restrições e tipo da função objetivo, para cada um dos problemas.

Problemas	Variáveis	Restrições	Tipo de Restrições	Tipo da Função Objetivo
CB2	3	3	3 não lineares	linear
CB3	3	3	3 não lineares	linear
CHACONN1	3	3	3 não lineares	linear
CHACONN2	3	3	3 não lineares	linear
COSHFUN	10	3	3 não lineares	linear
DEMYMALO	3	3	1 não linear 2 lineares	linear
DIPIGRI	7	4	4 não lineares	não linear
EXPFITA	5	22	22 lineares	não linear
EXPFITB	5	102	102 lineares	não linear
GIGOMEZ1	3	3	1 não linear 2 lineares	linear
HAIFAS	13	9	9 não lineares	linear
HALDMADS	6	42	42 não lineares	linear
HS10	2	1	1 não linear	linear
HS11	2	1	1 não linear	não linear
HS12	2	1	1 não linear	não linear
HS22	2	2	1 não linear 1 linear	não linear
HS29	3	1	1 não linear	não linear
HS43	4	3	3 não lineares	não linear
HS88	2	1	1 não linear	não linear
HS89	3	1	1 não linear	não linear
HS90	4	1	1 não linear	não linear
HS91	5	1	1 não linear	não linear
HS92	6	1	1 não linear	não linear
HS100	7	4	4 não lineares	não linear
HS100MOD	7	4	4 não lineares	não linear
HS113	10	8	5 não lineares 3 lineares	não linear

Tabela 6.1: Problemas do CUTE

continua...

Tabela 6.1 - Problemas do CUTE (continuação)

Problemas	Variáveis	Restrições	Tipo de Restrições	Tipo da Função Objetivo
HS268	5	5	5 lineares	não linear
KIWCRESC	3	2	2 não lineares	linear
LISWET1	103	100	100 lineares	não linear
LISWET2	103	100	100 lineares	não linear
LISWET3	103	100	100 lineares	não linear
LISWET4	103	100	100 lineares	não linear
LISWET5	103	100	100 lineares	não linear
LISWET6	103	100	100 lineares	não linear
LISWET10	103	100	100 lineares	não linear
MADSEN	3	6	6 não lineares	linear
MAKELA1	3	2	1 não linear 1 linear	linear
MAKELA2	3	3	3 não lineares	linear
MAKELA3	21	20	20 não lineares	linear
MAKELA4	21	40	40 lineares	linear
MIFFLIN1	3	2	1 não linear 1 linear	linear
MIFFLIN2	3	2	2 não lineares	linear
MINMAXBD	5	20	20 não lineares	linear
MINMAXRB	3	4	2 não lineares 2 lineares	linear
OET1	3	6	6 lineares	não linear
OET2	3	202	202 não lineares	linear
OET3	4	6	6 lineares	linear
OET3	4	202	202 lineares	linear
OET4	4	6	6 lineares	linear
OET5	5	6	6 não lineares	linear
OET6	5	6	6 não lineares	linear
OET7	7	6	6 não lineares	linear
PENTAGON	6	15	15 lineares	não linear
POLAK1	3	2	2 não lineares	linear
POLAK2	11	2	2 não lineares	linear
POLAK3	12	10	10 não lineares	linear
POLAK4	3	3	3 não lineares	linear
POLAK5	3	2	2 não lineares	linear
POLAK6	5	4	4 não lineares	linear
POWELL20	10	10	10 lineares	não linear
PT	2	501	501 lineares	linear
PT	2	3	3 lineares	linear
PT	2	101	101 lineares	linear
ROSENMMX	5	4	4 não lineares	linear
SIPOW1	2	20	20 lineares	linear
SIPOW1	2	100	100 lineares	linear
SIPOW1	2	500	500 lineares	linear

continua...

Tabela 6.1 - Problemas do CUTE (continuação)

Problemas	Variáveis	Restrições	Tipo de Restrições	Tipo da Função Objetivo
SIPOW1M	2	20	20 lineares	linear
SIPOW1M	2	100	100 lineares	linear
SIPOW1M	2	500	500 lineares	linear
SIPOW2	2	20	20 lineares	linear
SIPOW2M	2	20	20 lineares	linear
SIPOW2M	2	100	100 lineares	linear
SIPOW2M	2	500	500 lineares	linear
TFI1	3	11	11 não lineares	não linear
TFI1	3	51	51 não lineares	não linear
TFI2	3	11	11 lineares	linear
TFI2	3	101	101 lineares	linear
TFI3	3	11	11 lineares	não linear
TFI3	3	51	51 lineares	não linear
TFI3	3	101	101 lineares	não linear
WOMFLET	3	3	3 não lineares	linear

Os 82 problemas testados e listados na tabela (6.1) são bastante variados. A maioria é não convexo. Alguns são lineares, outros não lineares e outros são mistos.

6.1 Resultados Numéricos

Os resultados dos testes estão apresentados na tabela (6.2), cujos dados são especificados na primeira linha de cada coluna, ou seja:

- **Problema** - Nome do problema definido pela coleção CUTE.
- **Penalidade** - Tipo da penalidade e o tipo da função θ :
 1. quadr - é referente à quadrática (6.3)
 2. m2b - é referente à barreira logarítmica modificada (6.4)
- **it.int.** - Número de iterações executadas pelo algoritmo de Região de Confiança.
- **it.ext.** - Número de iterações do Lagrangeano Aumentado, ou seja, número de vezes que resolvemos o subproblema formado pela função Lagrangeano Aumentado.
- **função** - Número de avaliações de funções, incluindo todas, função objetivo e restrições.

- **grad.** - Número de avaliações dos gradientes, da função objetivo mais os das restrições.
- **L. A.** - Número de avaliações da função Lagrangeano Aumentado.
- **f. ótimo** - Valor ótimo da função objetivo encontrado pelo algoritmo.

Problema	Penalidade		it.int.	it.ext.	função	grad.	L. A.	f. ótimo
CB2	quadr	1	38	18	224	224	168	1.9522
		2	41	17	232	220	168	1.9522
	m2b	1	47	19	264	264	198	1.9522
		2	29	9	152	152	114	1.9522
CB3	quadr	1	26	11	148	148	111	2.0000
		2	29	9	152	136	106	2.0000
	m2b	1	35	11	184	184	138	2.0000
		2	31	9	160	160	120	2.0000
CHACONN1	quadr	1	45	18	252	252	189	1.9522
		2	44	17	244	232	177	1.9522
	m2b	1	34	14	192	184	140	1.9522
		2	20	5	100	92	71	1.9522
CHACONN2	quadr	1	15	5	80	80	60	2.0000
		2	13	4	68	68	51	2.0000
	m2b	1	28	6	136	132	100	2.0000
		2	32	5	148	128	101	2.0000
COSHFUN	quadr	1	17	4	84	76	59	-0.6614
		2	22	3	100	84	67	-0.6614
	m2b	1	20	5	100	92	71	0.6614
		2	29	4	132	100	83	-0.6614
DEMYMALO	quadr	1	8	3	44	44	33	-3.0000
		2	9	3	48	48	36	-3.0000
	m2b	1	14	4	72	72	54	-3.0000
		2	15	4	76	72	55	-3.0000
DIPIGRI	quadr	1	73	17	450	400	250	680.6301
		2	121	40	805	755	463	683.1228
	m2b	1	32	11	215	205	125	680.6301
		2	27	6	165	155	95	680.6301

Tabela 6.2: Resultados Numéricos

continua...

Tabela 6.2 - Resultados Numéricos (continuação)

Problema	Penalidade		it.int.	it.ext.	função	grad.	L. A.	f. ótimo
EXPFITA	quadr	1	56	40	2208	1242	204	4.4913e-4
		2	56	40	2208	1242	204	4.4913e-4
	m2b	1	56	40	2208	1242	204	4.4913e-4
		2	56	40	2208	1242	204	4.4913e-4
EXPFITB	quadr	1	55	40	9785	5562	203	0.0017
		2	55	40	9785	5562	203	0.0017
	m2b	1	55	40	9785	5562	203	0.0017
		2	55	40	9785	5562	203	0.0017
GIGOMEZ1	quadr	1	11	5	64	64	48	-3.0000
		2	11	5	64	64	48	-3.0000
	m2b	1	13	4	68	68	51	-3.0000
		2	13	3	64	64	48	-3.0000
HAIFAS	quadr	1	41	18	590	590	177	-0.4500
		2	79	40	1190	1140	347	-0.4500
	m2b	1	55	19	740	740	222	-0.4500
		2	39	10	490	460	141	-0.4500
HALDMADS	quadr	1	70	21	3913	3698	262	0.0341
		2	70	40	4730	4257	308	1.1468e-04
	m2b	1	54	20	3182	2838	206	1.2914e-04
		2	73	14	3741	3182	235	0.0333
HS10	quadr	1	23	6	58	58	87	-1.0000
		2	23	6	58	58	87	-1.0000
	m2b	1	28	6	68	68	102	-1.0000
		2	30	6	72	68	104	-1.0000
HS11	quadr	1	15	5	40	40	60	-8.4985
		2	17	6	46	46	69	-8.4985
	m2b	1	17	5	44	44	66	-8.4985
		2	17	5	44	44	66	-8.4985
HS100	quadr	1	76	20	480	430	268	680.6301
		2	121	40	805	755	463	683.1228
	m2b	1	34	13	235	225	137	680.6301
		2	28	7	175	165	101	680.6301
HS100MOD	quadr	1	447	29	2380	780	788	678.6796
		2	173	40	1065	960	597	697.0206
	m2b	1	221	11	1160	240	328	678.6796
		2	20	4	120	95	62	678.6796
HS12	quadr	1	14	4	36	36	54	-30.0000
		2	14	4	36	36	54	-30.0000
	m2b	1	16	4	40	38	58	-30.0000
		2	18	4	44	40	62	-30.0000

continua...

Tabela 6.2 - Resultados Numéricos (continuação)

Problema	Penalidade	it.int.	it.ext.	função	grad.	L. A.	f. ótimo	
HS22	quadr	1	15	7	66	66	66	1.0000
		2	14	6	60	60	60	1.0000
	m2b	1	19	7	78	78	78	1.0000
		2	19	6	75	75	75	1.0000
HS29	quadr	1	18	4	44	38	60	-22.6274
		2	18	4	44	38	60	-22.6274
	m2b	1	16	4	40	38	58	-22.6274
		2	16	4	40	38	58	-22.6274
HS43	quadr	1	32	13	180	176	133	-44.0000
		2	72	40	488	444	334	-44.0000
	m2b	1	39	14	212	204	155	-44.0000
		2	22	6	112	104	80	-44.0000
HS88	quadr	1	31	7	76	76	114	1.3627
		2	45	13	116	116	174	1.3627
	m2b	1	35	7	84	84	126	1.3627
		2	45	13	116	116	174	1.3627
HS89	quadr	1	33	7	80	78	118	1.3627
		2	45	13	116	114	172	1.3627
	m2b	1	35	7	84	82	124	1.3627
		2	46	13	118	116	175	1.3627
HS90	quadr	1	45	7	104	92	144	1.3627
		2	56	12	136	124	192	1.3627
	m2b	1	43	7	100	92	142	1.3627
		2	46	12	116	106	164	1.3627
HS91	quadr	1	45	7	104	92	144	1.3627
		2	55	13	136	124	192	1.3627
	m2b	1	46	7	106	92	145	1.3627
		2	55	13	136	122	190	1.3627
HS92	quadr	1	45	11	112	104	160	1.3627
		2	71	17	176	152	240	1.3627
	m2b	1	47	11	116	108	166	1.3627
		2	78	17	190	162	257	1.3627
HS113	quadr	1	52	19	639	639	213	24.3062
		2	102	40	1278	1278	426	24.2106
	m2b	1	41	17	522	522	174	24.3062
		2	32	8	360	333	114	24.3062
HS268	quadr	1	21	19	240	240	120	4.5940e-07
		2	42	40	492	492	246	1.1508e-02
	m2b	1	35	18	318	318	159	3.0586e-07
		2	24	9	198	198	99	9.0957e-08

continua...

Tabela 6.2 - Resultados Numéricos (continuação)

Problema	Penalidade	it.int.	it.ext.	função	grad.	L. A.	f. ótimo	
KIWCRES	quadr	1	21	4	75	69	71	2.8682e-08
		2	20	3	69	63	65	-1.1442e-08
	m2b	1	28	4	96	81	86	-4.6565e-08
		2	30	4	102	81	88	-6.8590e-09
LISWET1	quadr	1	17	15	3232	3232	96	0.2475
		2	42	40	8282	8282	246	0.2477
	m2b	1	51	14	6565	6565	195	0.2475
		2	99	14	11413	9090	293	0.2475
LISWET2	quadr	1	18	16	3434	3434	102	0.2530
		2	42	40	8282	8282	246	0.2504
	m2b	1	50	15	6565	6565	195	0.2530
		2	67	16	8383	7979	241	0.2530
LISWET3	quadr	1	20	18	3838	3838	114	0.2530
		2	42	40	8282	8282	246	0.2504
	m2b	1	58	18	7676	7676	228	0.2530
		2	76	19	9595	8989	273	0.2530
LISWET4	quadr	1	17	15	3232	3232	96	0.2513
		2	42	40	8282	8282	246	0.2503
	m2b	1	54	16	7070	7070	210	0.2513
		2	73	15	8888	8181	250	0.2513
LISWET5	quadr	1	19	15	3434	3434	102	0.2520
		2	44	40	8484	8484	252	0.2504
	m2b	1	50	14	6464	6464	192	0.2520
		2	66	14	8080	7676	232	0.2520
LISWET6	quadr	1	17	15	3232	3232	96	0.2540
		2	42	40	8282	8282	246	0.2504
	m2b	1	50	15	6565	6565	195	0.2540
		2	70	16	8686	8282	250	0.2540
LISWET10	quadr	1	15	13	2828	2828	84	0.2508
		2	42	40	8282	8282	246	0.2504
	m2b	1	51	14	6565	6565	195	0.2508
		2	70	12	8282	7272	226	0.2508
MADSEN	quadr	1	54	19	511	511	219	0.6164
		2	87	40	889	882	379	0.6127
	m2b	1	44	16	420	413	178	0.6164
		2	28	7	245	238	103	0.6164
MAKELA1	quadr	1	18	7	75	72	73	-1.4142
		2	16	6	66	63	64	-1.4142
	m2b	1	23	7	90	87	88	-1.4142
		2	21	6	81	78	79	-1.4142

continua...

Tabela 6.2 - Resultados Numéricos (continuação)

Problema	Penalidade		it.int.	it.ext.	função	grad.	L. A.	f. ótimo
MAKELA2	quadr	1	61	23	336	336	252	7.2000
		2	82	40	488	468	356	7.2000
	m2b	1	39	12	204	192	147	7.2000
		2	24	5	116	104	81	7.2000
MAKELA3	quadr	1	38	2	840	693	106	1.3130e-06
		2	39	2	861	714	109	-5.8202e-08
	m2b	1	22	2	504	483	70	7.9882e-08
		2	29	3	672	546	84	1.2108e-06
MAKELA4	quadr	1	7	2	369	369	27	1.4544e-14
		2	7	2	369	369	27	5.5511e-16
	m2b	1	19	2	861	820	61	3.1211e-07
		2	19	2	861	820	61	7.8029e-09
MIFFLIN1	quadr	1	6	3	27	27	27	-1.0000
		2	11	3	42	39	40	-1.0000
	m2b	1	5	2	21	21	21	-1.0000
		2	5	2	21	21	21	-1.0000
MIFFLIN2	quadr	1	23	7	90	90	90	-1.0000
		2	20	5	75	75	75	-1.0000
	m2b	1	27	7	102	99	100	-1.0000
		2	28	5	99	84	89	-1.0000
MINMAXBD	quadr	1	236	33	5649	4536	701	115.7064
		2	272	40	6552	5250	812	113.1018
	m2b	1	54	15	1449	1281	191	115.7064
		2	42	5	987	756	119	115.7064
MINMAXRB	quadr	1	29	3	160	135	86	2.5437e-13
		2	30	3	165	140	89	1.4775e-14
	m2b	1	35	3	190	180	110	6.3945e-08
		2	36	3	195	185	113	1.5986e-08
OET1	quadr	1	21	19	280	280	120	0.4038
		2	29	27	392	392	168	0.4038
	m2b	1	33	16	343	343	147	0.4038
		2	24	7	217	203	89	0.4038
OET2	quadr	1	36	29	13195	13195	195	0.5382
		2	47	40	17661	17661	261	0.4956
	m2b	1	55	19	15022	15022	222	0.5382
		2	60	15	15225	14413	217	0.5382
OET3	quadr	1	5	2	49	49	21	-5.3291e-15
		2	5	2	49	49	21	-1.7763e-15
	m2b	1	15	4	133	133	57	3.1701e-08
		2	15	2	119	119	51	9.1369e-08

continua...

Tabela 6.2 - Resultados Numéricos (continuação)

Problema	Penalidade	it.int.	it.ext.	função	grad.	L. A.	f. ótimo	
OET3 m=202	quadr	1	28	27	11165	11165	165	0.0045
		2	41	40	16443	16443	243	0.0040
	m2b	1	47	18	13195	12992	193	0.0045
		2	60	15	15225	13398	207	0.0045
OET4	quadr	1	19	2	147	98	49	-2.1094e-15
		2	19	2	147	98	49	-1.1102e-16
	m2b	1	30	3	231	161	79	2.0366e-06
		2	30	2	224	154	76	-4.0787e-10
OET5	quadr	1	8	2	70	70	30	-8.8818e-15
		2	10	4	98	98	42	-3.8399e-09
	m2b	1	17	4	147	140	61	2.1411e-08
		2	19	2	147	126	57	-3.5522e-08
OET6	quadr	1	30	4	238	203	92	-2.2226e-18
		2	30	4	238	203	92	-7.8978e-17
	m2b	1	34	5	273	224	103	1.6933e-09
		2	33	5	266	210	98	1.7054e-10
OET7	quadr	1	34	4	266	203	96	-5.3333e-08
		2	34	4	266	203	96	-1.2282e-09
	m2b	1	35	5	280	252	112	5.7360e-08
		2	38	4	294	252	114	3.9310e-09
PENTAGON	quadr	1	150	40	3040	2640	520	2.1307e-07
		2	500	40	8640	6864	1398	2.5788e-63
	m2b	1	281	26	4912	3856	789	1.4621e-04
		2	138	12	2400	1936	392	1.3653e-04
POLAK1	quadr	1	17	2	57	54	55	2.7183
		2	18	2	60	57	58	2.7183
	m2b	1	21	2	69	63	65	2.7183
		2	24	4	84	75	78	2.7183
POLAK2	quadr	1	10	2	36	36	36	54.5982
		2	26	2	84	63	70	54.5982
	m2b	1	15	2	51	48	49	54.5982
		2	18	4	66	60	62	54.5982
POLAK3	quadr	1	149	21	1870	1540	450	5.9330
		2	186	40	2486	2068	602	5.9215
	m2b	1	53	16	759	759	207	5.9330
		2	42	10	572	572	156	5.9330
POLAK4	quadr	1	11	6	68	68	51	1.5707e-07
		2	8	2	40	40	30	5.9912e-12
	m2b	1	44	9	212	208	157	2.9000e-07
		2	60	3	252	168	147	1.5966e-08

continua...

Tabela 6.2 - Resultados Numéricos (continuação)

Problema	Penalidade	it.int.	it.ext.	função	grad.	L. A.	f. ótimo	
POLAK5	quadr	1	98	3	303	240	261	50.0000
		2	97	3	300	243	262	50.0000
	m2b	1	73	3	228	201	210	50.0000
		2	73	3	228	201	210	50.0000
POLAK6	quadr	1	341	24	1825	1290	881	-44.0000
		2	398	40	2190	1545	1056	-44.0000
	m2b	1	48	14	310	295	180	-44.0000
		2	32	5	185	160	101	-44.0000
POWELL20	quadr	1	15	13	308	308	84	57.8125
		2	34	32	726	726	198	57.8125
	m2b	1	26	11	407	407	111	57.8125
		2	27	11	418	418	114	57.8125
PT	quadr	1	56	35	45682	43674	265	0.1784
		2	47	40	43674	43674	261	0.1689
	m2b	1	62	21	41666	41164	247	0.1784
		2	89	18	53714	46184	291	0.1784
PT m=3	quadr	1	15	15	120	120	90	0.1429
		2	11	11	88	88	66	0.1429
	m2b	1	31	16	188	184	139	0.1429
		2	19	7	104	104	78	0.1429
PT m=101	quadr	1	36	31	6834	6834	201	0.1784
		2	45	40	8670	8670	255	0.1680
	m2b	1	48	16	6528	6528	192	0.1784
		2	61	14	7650	7038	213	0.1784
ROSENMMX	quadr	1	93	21	570	520	322	-44.0000
		2	122	40	810	735	456	-44.0000
	m2b	1	47	16	315	315	189	-44.0000
		2	32	7	195	180	111	-44.0000
SIPOW1	quadr	1	20	19	819	819	117	-1.0000
		2	41	40	1701	1701	243	-1.0701
	m2b	1	38	19	1197	1197	171	-1.0000
		2	31	11	882	840	122	-1.0000
SIPOW1 m=100	quadr	1	21	21	4242	4242	126	-1.0000
		2	40	40	8080	8080	240	-1.0756
	m2b	1	42	19	6161	6161	183	-1.0000
		2	35	14	4949	4848	145	-1.0000
SIPOW1 m=500	quadr	1	25	25	25050	25050	150	-1.0000
		2	40	40	40080	40080	240	-1.0799
	m2b	1	46	19	32565	32565	195	-1.0000
		2	47	16	31563	29559	181	-1.0000

continua...

Tabela 6.2 - Resultados Numéricos (continuação)

Problema	Penalidade		it.int.	it.ext.	função	grad.	L. A.	f. ótimo
SIPOW1M	quadr	1	17	17	714	714	102	-1.0125
		2	40	40	1680	1680	240	-1.0736
	m2b	1	32	16	1008	1008	144	-1.0125
		2	18	8	546	525	76	-1.0125
SIPOW1M m=100	quadr	1	19	19	3838	3838	114	-1.0005
		2	40	40	8080	8080	240	-1.0756
	m2b	1	42	19	6161	6161	183	-1.0005
		2	32	13	4545	4444	133	-1.0005
SIPOW1M m=500	quadr	1	23	23	23046	23046	138	-1.0000
		2	40	40	40080	40080	240	-1.0990
	m2b	1	40	16	28056	28056	168	-1.0000
		2	40	13	26553	24549	151	-1.0000
SIPOW2	quadr	1	19	19	798	798	114	-1.0515
		2	40	40	1680	1680	240	-1.0717
	m2b	1	37	16	1113	1113	159	-1.0515
		2	22	7	609	588	85	-1.0515
SIPOW2M	quadr	1	19	19	798	798	114	-1.0000
		2	40	40	1680	168/0	240	-1.0672
	m2b	1	32	16	1008	1008	144	-1.0000
		2	22	8	630	609	88	-1.0000
SIPOW2M m=100	quadr	1	25	25	5050	5050	150	-1.0000
		2	40	40	1680	1680	240	-1.0672
	m2b	1	45	16	6161	6161	183	-1.0000
		2	34	10	4444	4141	126	-1.0000
SIPOW2M m=500	quadr	1	33	32	32565	32565	195	-1.0000
		2	41	40	40581	40581	243	-1.0738
	m2b	1	47	14	30561	30561	183	-1.0000
		2	37	10	23547	22044	135	-1.0000
TFI1	quadr	1	70	21	1092	972	253	5.3347
		2	106	40	1752	1608	414	5.3172
	m2b	1	45	16	732	696	177	5.3347
		2	34	7	492	396	107	5.3347
TFI1 m=51	quadr	1	1574	40	83928	25376	2590	3.0000
		2	1574	40	83928	25376	2590	3.0000
	m2b	1	206	19	11700	7592	517	5.3689
		2	48	9	2964	2444	151	5.3347
TFI2	quadr	1	16	15	372	372	93	0.6479
		2	41	40	972	972	243	0.6466
	m2b	1	38	16	648	624	158	0.6479
		2	40	10	600	528	138	0.6479

continua...

Tabela 6.2 - Resultados Numéricos (continuação)

Problema	Penalidade	it.int.	it.ext.	função	grad.	L. A.	f. ótimo	
TFI2 m=101	quadr	1	27	25	5304	5304	156	0.6490
		2	42	40	8364	8364	246	0.6456
	m2b	1	45	16	6222	6222	183	0.6490
		2	58	14	7344	6732	204	0.6490
TFI3	quadr	1	30	19	588	588	147	4.3011
		2	51	40	1092	1092	273	4.2965
	m2b	1	37	16	636	636	159	4.3011
		2	33	11	528	516	130	4.3011
TFI3 m=51	quadr	1	54	30	4368	4368	252	4.3011
		2	59	40	5148	5148	297	4.2959
	m2b	1	43	14	2964	2964	171	4.3011
		2	41	11	2704	2652	154	4.3011
TFI3 m=101	quadr	1	62	28	9180	9180	270	4.3012
		2	76	40	11832	11832	348	4.2929
	m2b	1	52	16	6936	6936	204	4.3012
		2	50	14	6528	6426	190	4.3012
WOMFLET	quadr	1	54	3	228	156	135	9.8533e-09
		2	52	3	220	148	129	-4.5187e-13
	m2b	1	45	4	196	140	119	-1.9394e-10
		2	40	4	176	120	104	9.0368e-17

Para simplificar a análise dos resultados, construiremos outras tabelas menores que darão informações específicas sobre alguns dados da mesma.

Começamos com a tabela (6.3), em que compararemos as penalidades **tipo 1** e **2** com as funções quadr e m2b.

	Penalidade	it.int.	it.ext.	função	grad	L.A.
tipo1	quadr	58	17	55	54	55
	m2b	22	39	25	23	25
	iguais	2	26	2	5	2
tipo2	quadr	37	10	28	25	27
	m2b	39	52	48	51	51
	iguais	6	20	6	6	4

Tabela 6.3: Comparação entre Quadrática e M2b

Dessa forma, na tabela (6.3) foi computado o número de vezes em que cada critério foi melhor, no sentido de que executou um menor número de iterações. Por exemplo, na primeira coluna da tabela (6.3) com penalidade **tipo 1** aparecem os números: 58, 22 e

2. Isso significa que a penalidade **tipo 1** com quadr resolveu 58 problemas executando menos iterações que a **tipo 1** com m2b. Da mesma forma, a penalidade **tipo 1** com m2b resolveu 22 problemas executando menos iterações que a **tipo 1** com quadr. Além disso, houve 2 empates, ambas executaram o mesmo número de iterações. O restante da tabela segue essa mesma linha.

Na tabela (6.3) tem-se que, com a penalidade **tipo 1** a quadrática foi superior a m2b em quase todos os critérios, exceto o critério de iteração externa. Porém, com a penalidade **tipo 2** a função m2b foi melhor em todos os critérios.

Como na tabela (6.3) a penalidade **tipo 1** foi melhor com θ quadrática e a penalidade **tipo 2** foi melhor com θ m2b, faremos na tabela (6.4) a comparação entre essas duas metodologias.

Penalidade		it.int.	it.ext.	função	grad.	L.A.
quadr	Tipo 1	56	13	45	44	44
m2b	Tipo 2	24	51	34	34	35
	iguais	2	18	3	4	3

Tabela 6.4: Comparação entre Quadrática Tipo1 e M2b Tipo 2

A função quadrática com penalidade **tipo 1** foi superior a função m2b com penalidade **tipo 2** em quase todos os critérios, exceto o critério de iteração externa.

Comentários sobre os testes numéricos

O principal objetivo dos testes numéricos é avaliar o desempenho do algoritmo de Lagrangeano Aumentado, utilizando a metodologia que estamos propondo. Pelos testes que fizemos e os resultados mostrados nas tabelas desta seção, vimos que o algoritmo de Lagrangeano Aumentado com penalidade **tipo 1** e θ quadrática proposta neste trabalho teve um desempenho superior, que as penalidades **tipo 1** e **2** formadas com θ sendo a m2b.

6.2 Resultados sobre Heurística da Atualização do Parâmetro de Penalidade

Para simplificar a análise dos resultados, construiremos apenas tabelas menores que darão informações específicas sobre alguns dados da mesma. Os números listados nas tabelas desta seção foram calculados da mesma forma que na tabela (6.2).

A notação para as funções θ especificadas na primeira coluna, são:

- quadr H - Algoritmo (5.3.3) com penalidade p **tipo 1** e **2**, θ quadrática com heurística;
- quadr - Algoritmo (5.3.1) com penalidade p **tipo 1** e **2** e θ quadrática;
- m2b - Algoritmo (5.3.2) com penalidade p **tipo 1** e **2** e θ m2b.

Começamos com a tabela (6.5), em que computaremos quais das duas funções quadr H e quadr tem um melhor desempenho tanto com as penalidades **tipo 1** como as de **tipo 2**. A explicação para os valores em cada coluna é a mesma que foi dada para a tabela (6.3).

Penalidade		it.int.	it.ext.	função	grad	L.A.
tipo1	quadr H	17	21	22	21	21
	quadr	16	11	14	15	15
	iguais	49	50	46	46	46
tipo2	quadr H	19	9	19	19	19
	quadr	15	5	15	15	15
	iguais	48	68	48	48	48

Tabela 6.5: Comparação entre as Quadráticas

Nas penalidades **tipo 1** e **2** a quadrática com heurística foi superior a quadrática em todos os critérios.

Na tabela (6.6) faremos a comparação entre as penalidades **tipo 1** e **2** quadrática heurística, como feito na tabela (6.4).

Penalidade		it.int.	it.ext.	função	grad.	L.A.
quadr. H	Tipo 1	52	41	53	51	52
quadr. H	Tipo 2	17	13	17	19	17
	iguais	13	28	12	12	13

Tabela 6.6: Comparação das Penalidades Tipo 1 e 2 Quadrática com Heurística

Com a heurística para a atualização do parâmetro de penalidade, o algoritmo com a penalidade **tipo 1** e θ quadrática foi superior ao algoritmo com penalidade **tipo 2** e θ quadrática.

Na tabela (6.7), computaremos quais das duas funções quadr H e m2b tem um melhor desempenho tanto com as penalidades **tipo 1** como as de **tipo 2**, como feito nas tabelas (6.3) e (6.5).

Penalidade		it.int.	it.ext.	função	grad	L.A.
tipo1	quadr H	53	23	54	54	54
	m2b	27	32	26	25	26
	iguais	2	27	2	3	2
tipo2	quadr H	37	13	28	26	28
	m2b	41	54	49	52	51
	iguais	4	15	5	4	3

Tabela 6.7: Comparação entre Quadrática com Heurística e M2b

Na penalidade **tipo 1** a quadrática foi superior a m2b em quase todos os critérios, exceto o critério de iteração externa. Porém com a penalidade **tipo 2** a função m2b foi melhor em todos os critérios.

Na tabela (6.8) faremos a comparação entre a penalidade **tipo 1** com quadrática e penalidade **tipo 2** com m2b, como feito nas tabelas (6.4) e (6.6).

Penalidade		it.int.	it.ext.	função	grad.	L.A.
quadr. H	Tipo 1	52	12	43	43	41
m2b	Tipo 2	26	51	35	36	38
	iguais	4	19	4	3	3

Tabela 6.8: Comparação entre Quadrática com Heurística Tipo 1 e M2b Tipo 2

Com a heurística para a atualização do parâmetro de penalidade, o algoritmo com a penalidade **tipo 1** e θ quadrática foi superior ao algoritmo com penalidade **tipo 2** e θ m2b.

Comentários sobre os testes numéricos da heurística

Pelos testes executados e os resultados mostrados nas tabelas desta seção, vê-se que o algoritmo de Lagrangeano Aumentado com a penalidade **tipo 1** e θ quadrática com heurística, proposta neste trabalho, teve um desempenho superior aos algoritmos de Lagrangeano Aumentado com as demais penalidades.

Também podemos notar que no algoritmo de Lagrangeano Aumentado com a penalidade **tipo 1** e θ quadrática geram-se subproblemas irrestritos mais fáceis de resolver que as demais penalidades, pois executa menos iterações internas conforme foi mostrado nas tabelas apresentadas neste capítulo.

Considerações Finais

Conclusão

A metodologia proposta aqui mostrou-se promissora. Com a implementação em Matlab dos algoritmos de Lagrangeano Aumentado verificou-se que, com a penalidade p **tipo 1** e função θ quadrática, os resultados numéricos são promissores. Entre as quadráticas, a usada com heurística para atualizar o parâmetro de penalidade foi superior à outra que não usou heurística.

No Capítulo 4, apresentamos um algoritmo com penalidade quadrática, propondo uma metodologia para contornar o problema dos multiplicadores serem negativos. Acreditamos que possa ser um bom algoritmo, pois os testes numéricos no Capítulo 6 foram promissores. Falta fazer alguns estudos sobre seu comportamento teórico, em relação à convergência.

Sugestões para Trabalhos Futuros

Abaixo são expostas algumas sugestões para trabalhos futuros nesta área:

- Estudos teóricos com relação à convergência da penalidade quadrática no algoritmo de Lagrangeano Aumentado;
- Testes computacionais com função objetivo quadrática e restrições lineares;
- Testes computacionais em Fortran com problemas de grande porte.

Bibliografia

- [1] D. P. Bertsekas. *Constrained Optimization and Lagrange Multipliers*. Academic Press, Ney York, 1992.
- [2] Y. Censor, S. Zenios. *The Proximal Minimization Algorithm with d-functions*. Journal Optmization Theory and Aplications, 73: pp.451-464, 1992.
- [3] G. Chen, M. Teboulle. *Convergence Analysis of a Proximal-like Optimization Algorithm using Bregman Functions*. SIAM J.Optimization Vol. 3, N.º3, pp.538-543, 1993.
- [4] J. E. Dennis, R. B. Schnabel. *Numerical Methods for Unconstraneid Optimization and Nonlinear Equations*. Englewood Clifts, Prentice Hall, 1983.
- [5] J. Eckstein. *Nonlinear Proximal Point Algorithms using Bregman Functions, with Applications to Convex Programming*. Mathematics of Operations Resaerchs, Vol. 18, N.º1, pp.:202-226, 1993.
- [6] Y. C. R. Espinoza. *Um Teorema de Equivalência entre Métodos Lagrangeano Aumentado e Algoritmos de Pontos Proximais*. Master's thesis, Universidade Federal de Santa Catarina, Santa Catarina, Brazil, 1998.
- [7] A. Friedlander. *Elementos de Programação Não-linear*, Editora UNICAP, 1994.
- [8] H. Frietzsche. *Programação Não-Linear Análise e Métodos*. São Paulo, Editora Edgard Blücher, 1977.
- [9] C. C. Gonzaga. *Algoritmos de Pontos Interiores para Programação Linear*. 17.ºColóquio Brasileiro de Matemática, Rio de Janeiro, IMPA, 1989.
- [10] C. C. Gonzaga. *Path Following Methods for Linear Programming*. SIAM Review, 34(2): pp.167-227, 1992.
- [11] J-B. Hiriart-Urruty, C. Lemarechal. *Convex Analysis and Minimization Algorithms I*. Springer-Verlag, New York, 1993.

- [12] J-B. Hiriart-Urruty, C. Lemarechal. *Convex Analysis and Minimization Algorithms II*. Springer-Verlag, New York, 1993.
- [13] A. R. Conn I. Bongartz, N. I. M. Gould and Ph. L. Toint. CUTE: *Constrained and Unconstrained Testing Environment*. ACM Transactions on Mathematical Software, 21: pp. 123-160, 1995.
- [14] A. Iusem. *Métodos de Ponto Proximal em Otimização*, 20.º Colóquio Brasileiro de Matemática, Rio de Janeiro, IMPA, 1995.
- [15] E. L. Lima. *Curso de Análise*, Rio de Janeiro, IMPA, 3.^a edição, Vol. 2, 1989.
- [16] J. M. Martínez, S. A. Santos. *Métodos Computacionais de Otimização*, 20.º Colóquio Brasileiro de Matemática, Rio de Janeiro, IMPA, 1995.
- [17] L.C. Mاتيoli. *Uma nova Metodologia para Construção de Funções de Penalização para Algoritmo de Lagrangeano Aumentado*, Tese doutorado, Universidade Federal de Santa Catarina, Florianópolis, Brasil, 2000.
- [18] M. Minoux. *Mathematical Programming*. John Wiley & Sons Ltd, 1986.
- [19] J. Nocedal, S. J. Wright. *Numerical Optimization*. New York: Springer-Verlag, 1999.
- [20] R. T. Rockafellar. *Augmented Lagrangians and Applications of the Proximal Point Algorithm in Convex Programming*. Mathematics of Operations Research, 1(2): pp 97-116, 1976.
- [21] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, New Jersey, 1970.
- [22] R. T. Rockafellar. *Monotone Operators and the Proximal Point Algorithm*. SIAM Journal on Control Optimization, 14(5): pp. 877-898, 1976.
- [23] R. T. Rockafellar. *The Multiplier Method of Hestenes and Powell Applied to Convex Programming*. Journal of Optimization Theory and Applications, 12: pp. 555-562, 1973.
- [24] M. Teboulle. *Entropic Proximal Mappings with Applications to Nonlinear Programming*. Mathematics of Operations Research, 17: pp. 97-116, 1992.
- [25] P. Tseng, D. Bertsekas. *On the Convergence of Exponential Multiplier Method for Convex Programming*. Mathematical Programming, 60: pp. 1-19, 1993.